KITS Workshop on “Machine Learning and Many-Body Physics”

Sparse modeling: how to solve the ill-posed problem

Graduate School of Information Sciences
Tohoku University

Masayuki OHZEKI

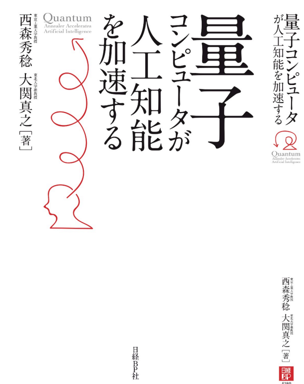
CREST Sparse Modeling ImPACT

QUANTUM ANNEALING MACHINE LEARNING

TOHOKU UNIVERSITY

Self Introduction

- Masayuki Ohzeki (大関真之)
  - Tohoku University [2016.10-]
  - Kyoto University [2010.05-2016.09]
    - Machine learning: Sparse modeling and Deep learning
  - Tokyo Institute of Technology [2008.10-2010.04]
    - Physics: Statistical Mechanics and Quantum annealing



QUANTUM ANNEALING

MACHINE LEARNING

Me in Japan… Cheating OHZEKI ?

- Mar. 27. 2015:NHK「Good morning Japan」

Recent advance
in news and media

Recent advance
in news and media

A. I. ?

Recent advance
in news and media

Inverse problem

$$y = f(\mathbf{x})$$

Recent advance
in news and media

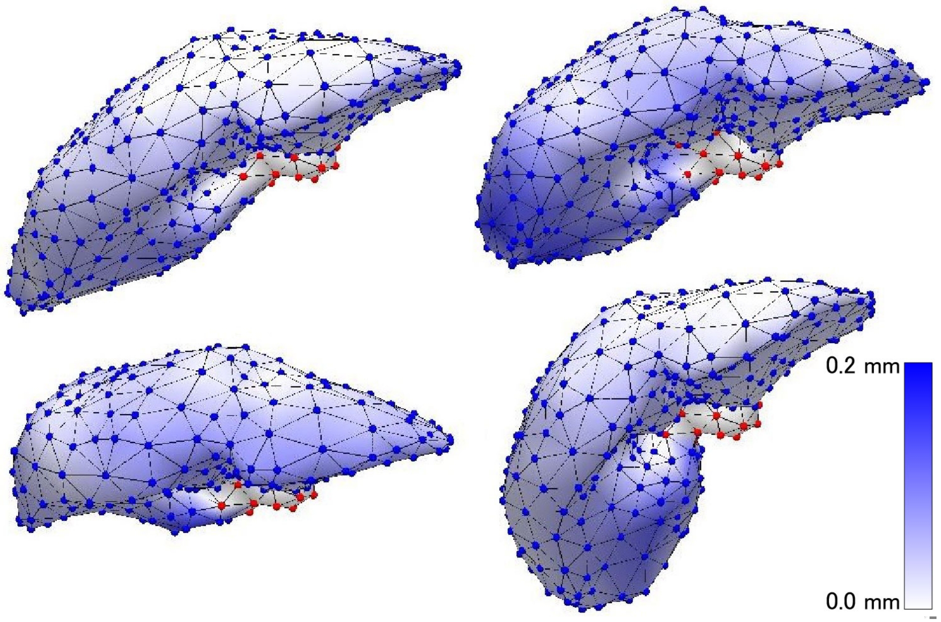
Inverse problem

$$y = f(\mathbf{x})$$

Estimation for deformation in body

Collaboration with U. Yamamoto, M. Kaneko, T. Matsuda (Kyoto Univ.) submitted

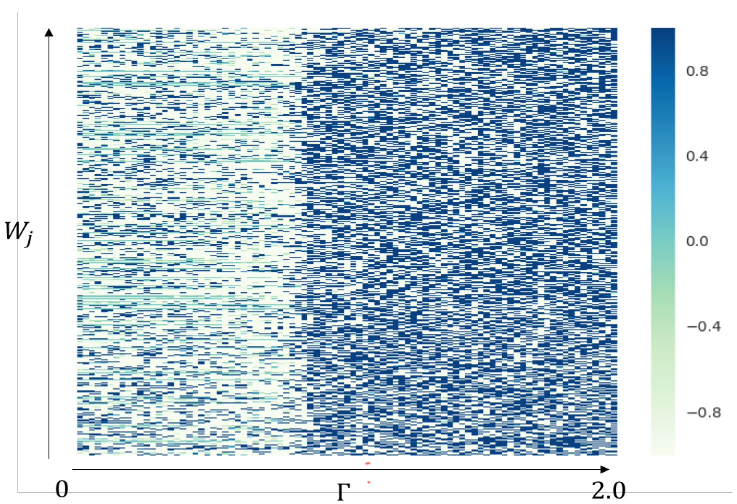
- Inference from a limited range of data
  - Learning the rule of the dynamics
  - Observation 9 points • unobserved 295 points
  - Inference of the deformation by pulling 4cm
  - (average of RMSE 0.035mm!)

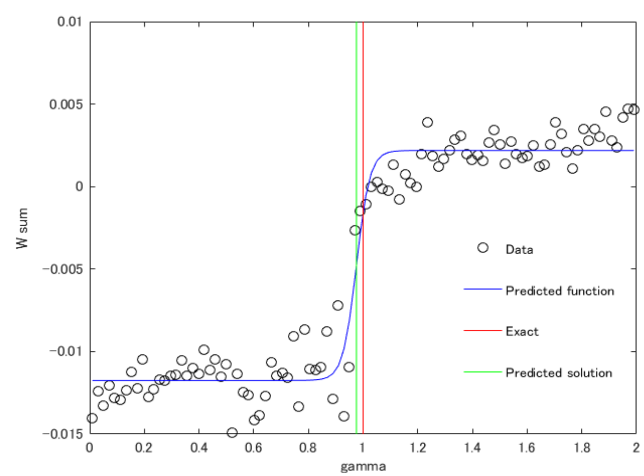


Detecting the phase transition

Collaboration with S. Arai (Tohoku Univ.) to appear soon

- Original work done by Tanaka and Tomiya (2017)
  - Learning from snapshots of the Ising model
  - Detection of the second-order phase transition from weights
- Extension of their work
  - Learning from snapshots of the XY/quantum Ising model
  - Detection of the Kosterlitz-Thouless transition
  - Detection of the second-order phase transition in quantum system





CREST Sparse Modeling ImPACT QUANTUM ANNEALING MACHINE LEARNING

From Purchase data, Customer service

Am○zon

From web experience, user service

Go◯gle

Recent advance
in news and media

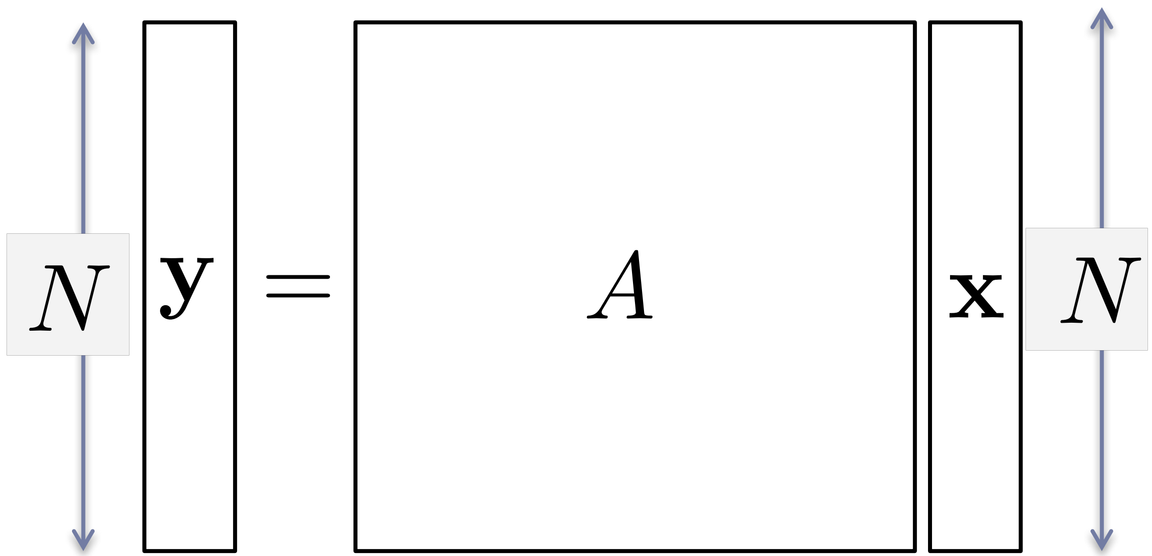
Inverse problem

$$y = f(\mathbf{x})$$

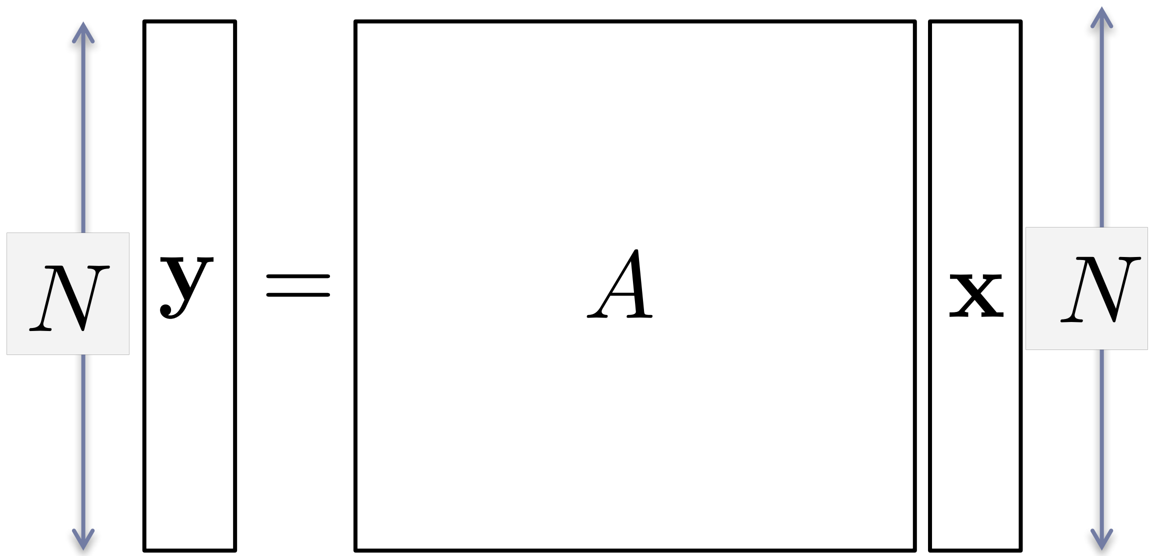
Recent advance
in news and media

Inverse problem

$$y = f(\mathbf{x})$$

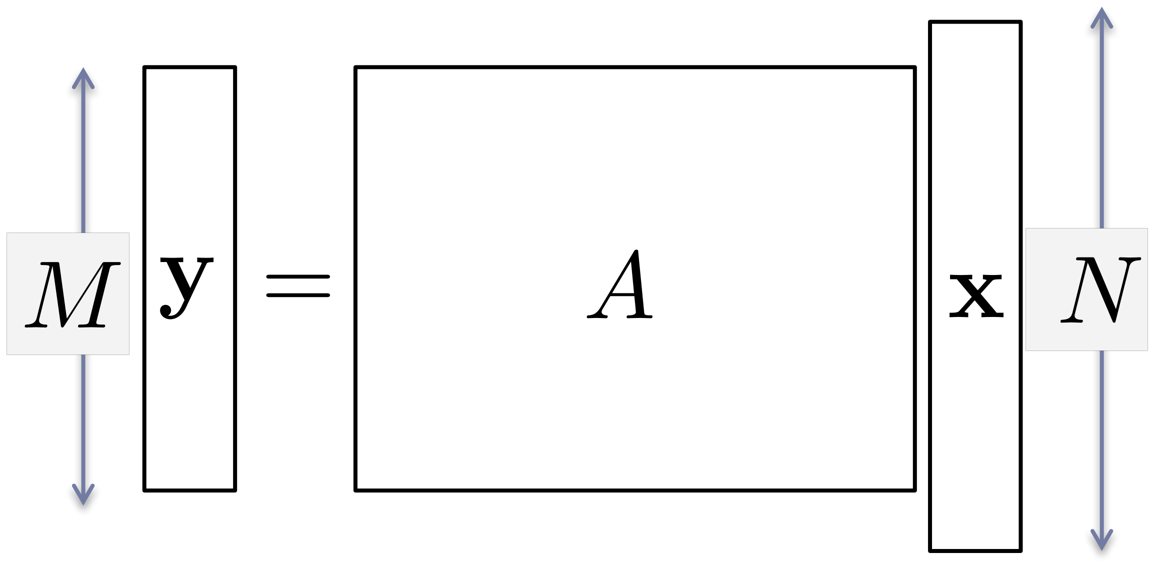


$$\mathbf{y} = A\mathbf{x}$$



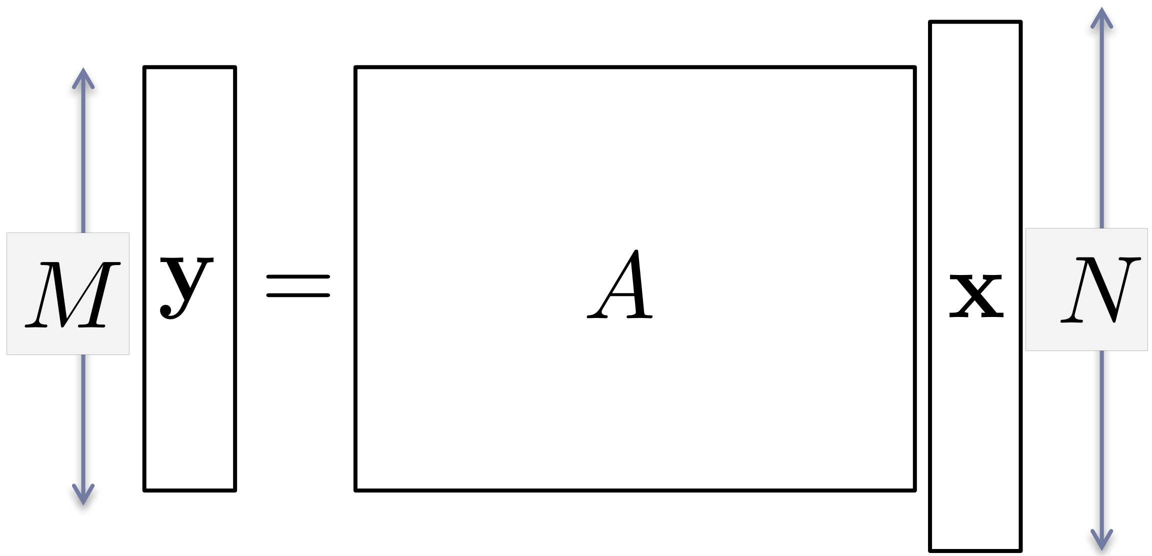
$$\mathbf{y} = A\mathbf{x}$$

Multiplication of inverse of A



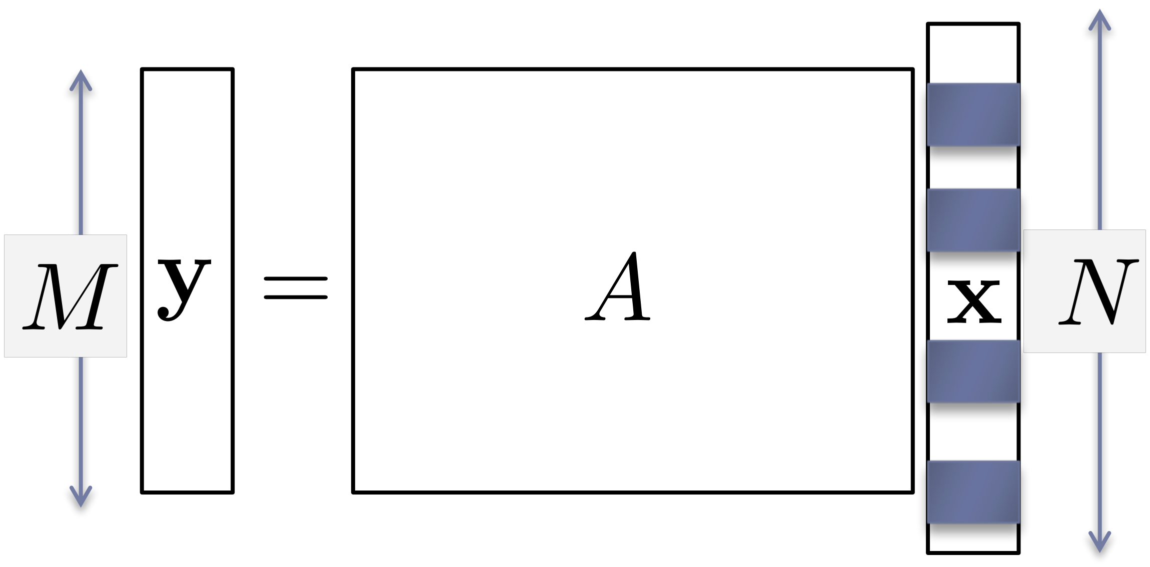
$$\mathbf{y} = A\mathbf{x}$$

In this case?



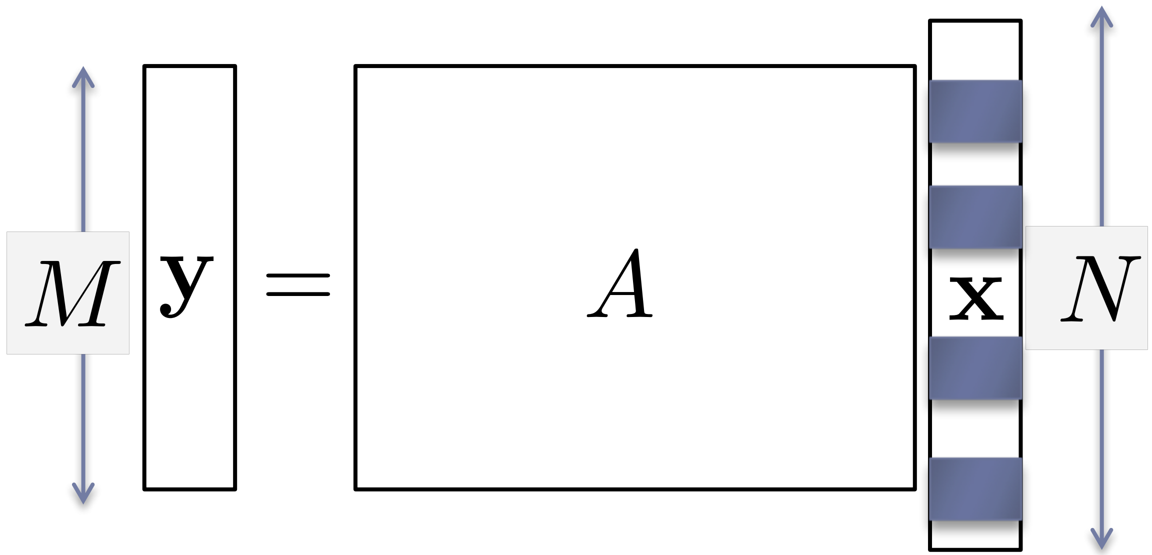
$$\mathbf{y} = A\mathbf{x}$$

No inverse matrix



$$\mathbf{y} = A\mathbf{x}$$

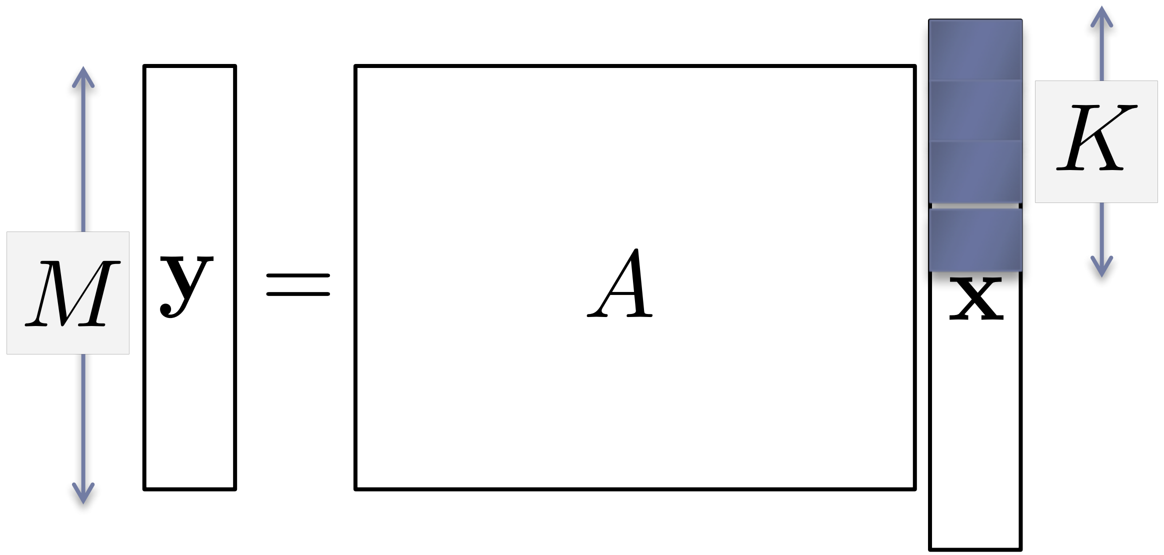
In the case with sparse solution



$$\mathbf{y} = A\mathbf{x}$$

In the case with sparse solution

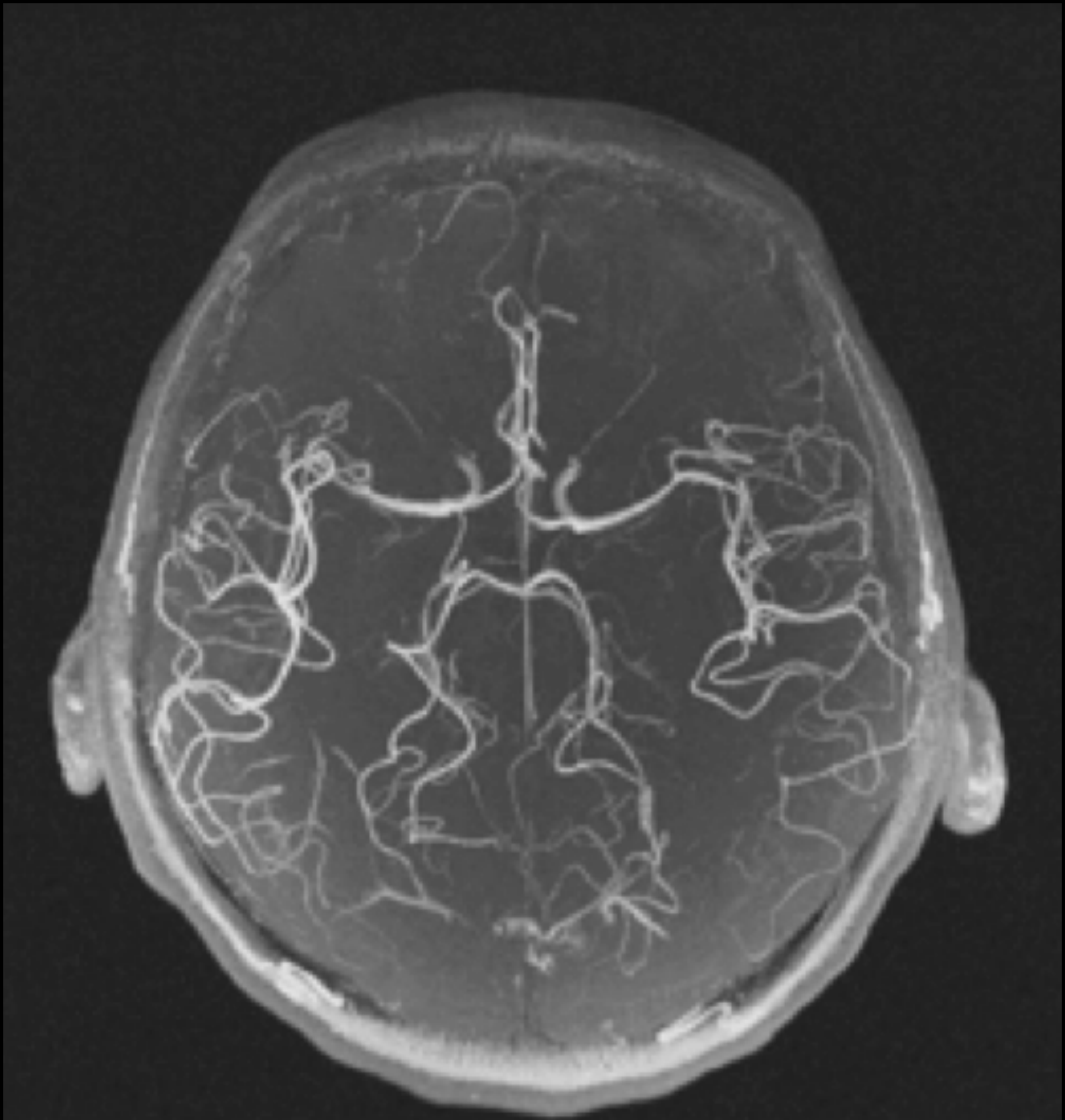
We can



$$\mathbf{y} = A\mathbf{x}$$

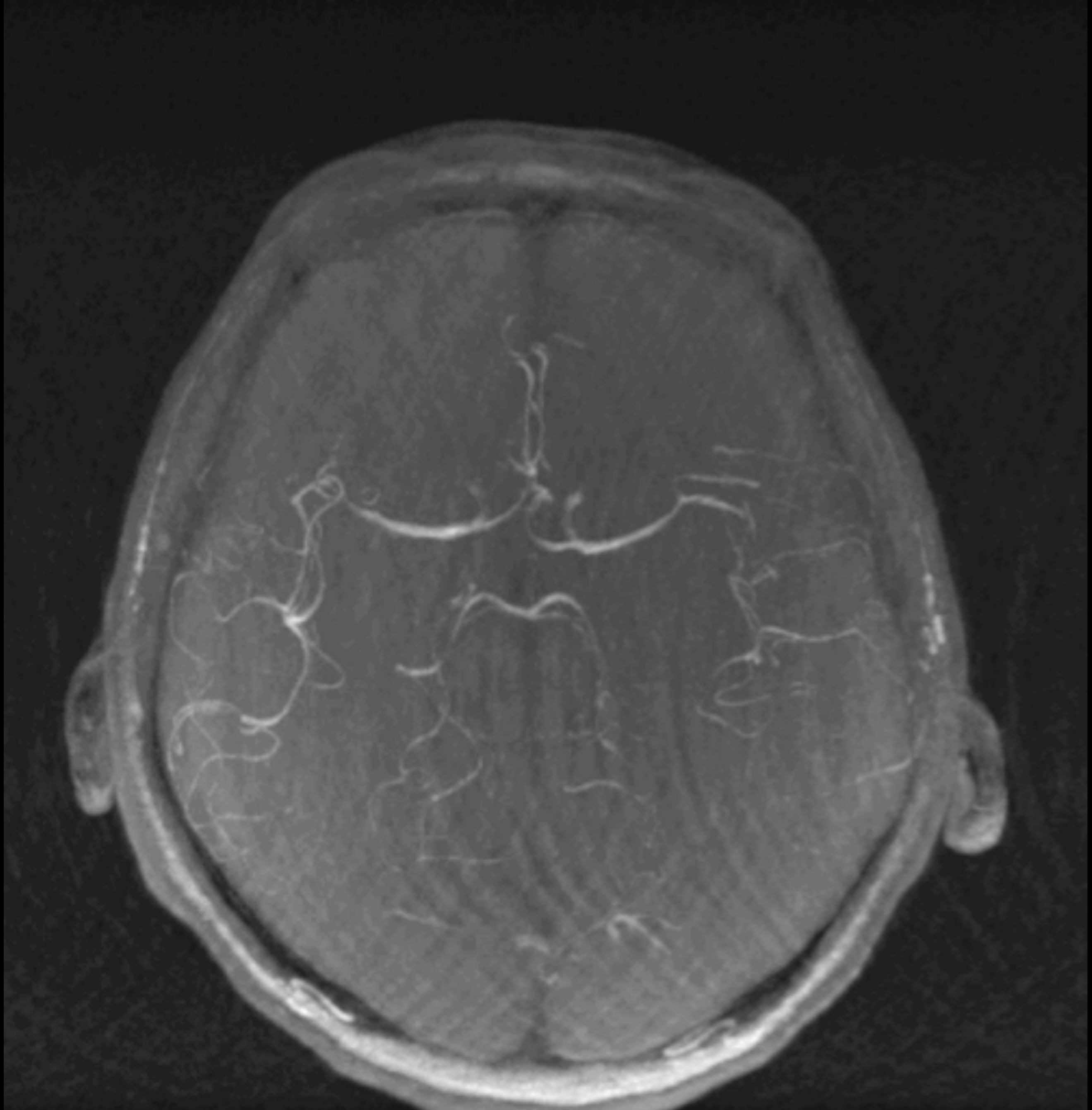
In the case with sparse solution

We can



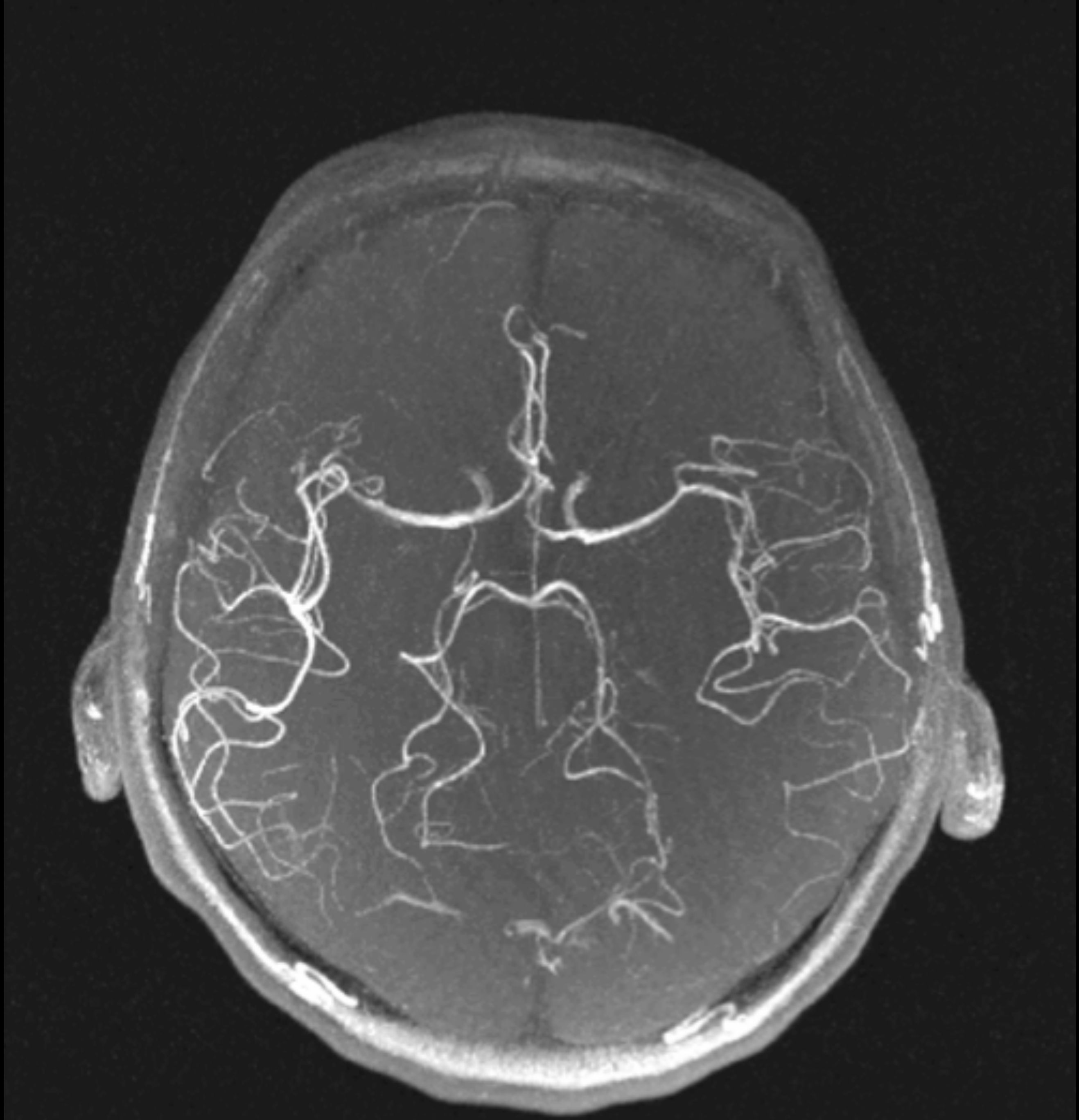
Collaboration with Kyoto University

Example of MRI



Collaboration with Kyoto University

Low amount of data



Collaboration with Kyoto University

Compressed sensing

Astrophysics and Sparse modeling

K. Akiyama, et al. Astrophys. J. (2017)

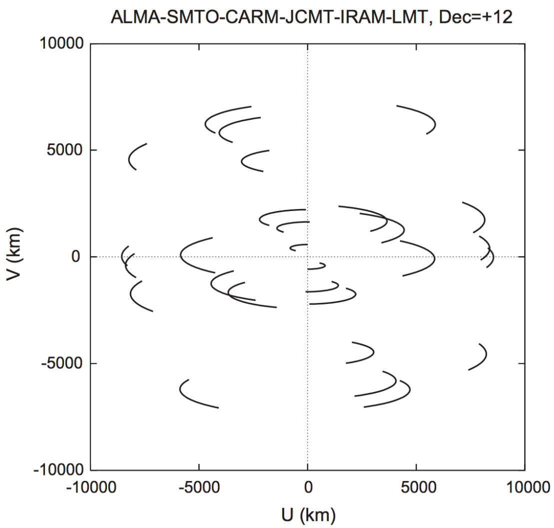
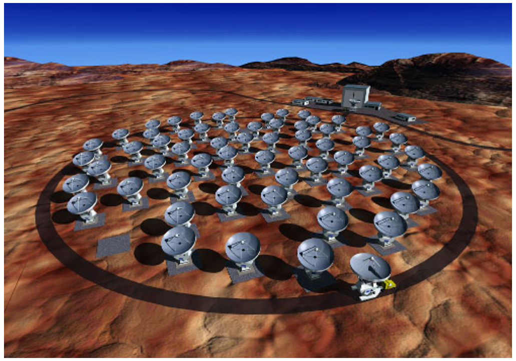


Fig. 5. Simulated UV coverage of M87 with six-station sub-mm VLBI array of EHT. Here it is assumed that observations are conducted at an elevation larger than 20° at each station.



CREST Sparse Modeling ImPACT QUANTUM ANNEALING MACHINE LEARNING

Astrophysics and Sparse modeling

K. Akiyama, et al. Astrophys. J. (2017)

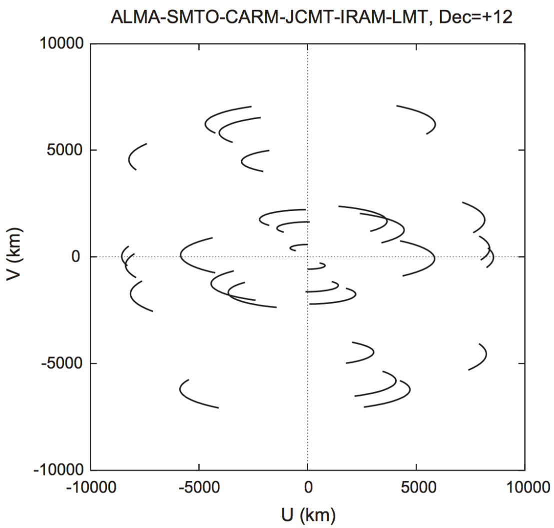
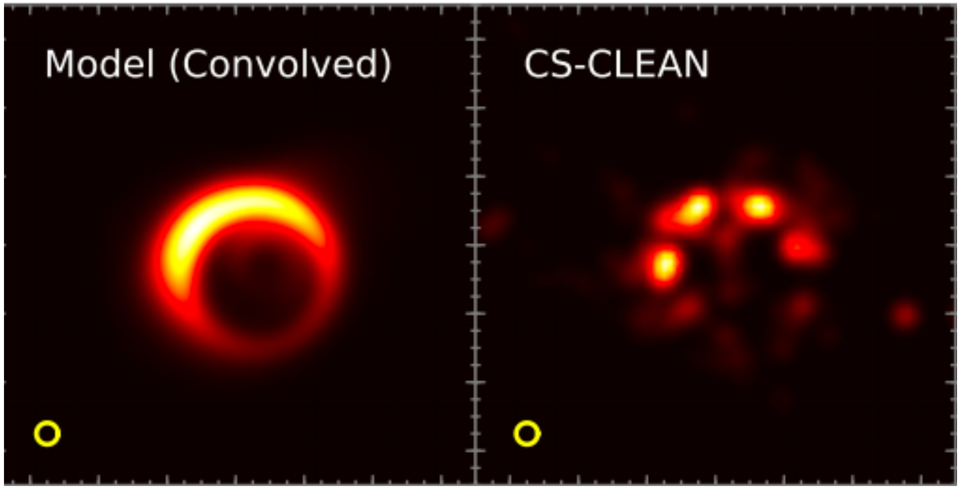
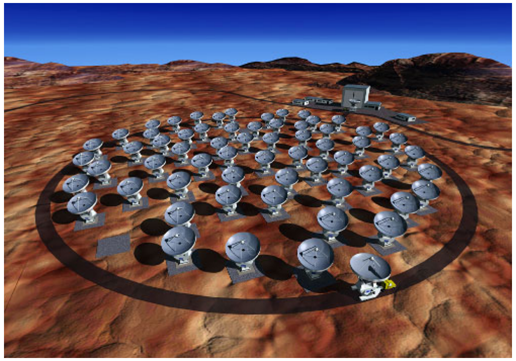


Fig. 5. Simulated UV coverage of M87 with six-station sub-mm VLBI array of EHT. Here it is assumed that observations are conducted at an elevation larger than 20° at each station.





CREST Sparse Modeling ImPACT QUANTUM ANNEALING MACHINE LEARNING

Astrophysics and Sparse modeling

K. Akiyama, et al. Astrophys. J. (2017)

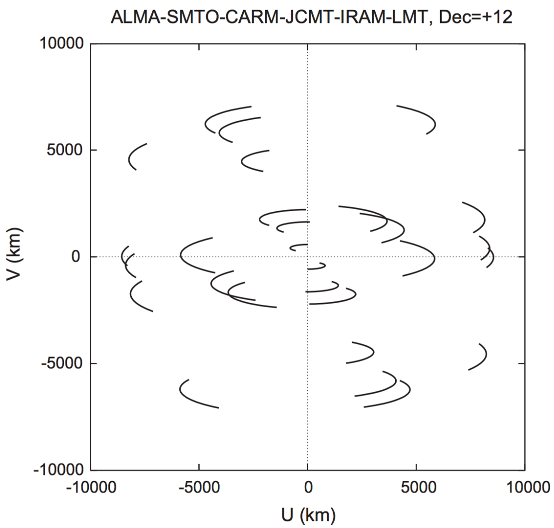
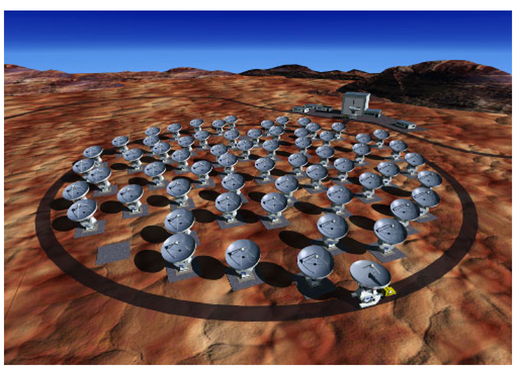
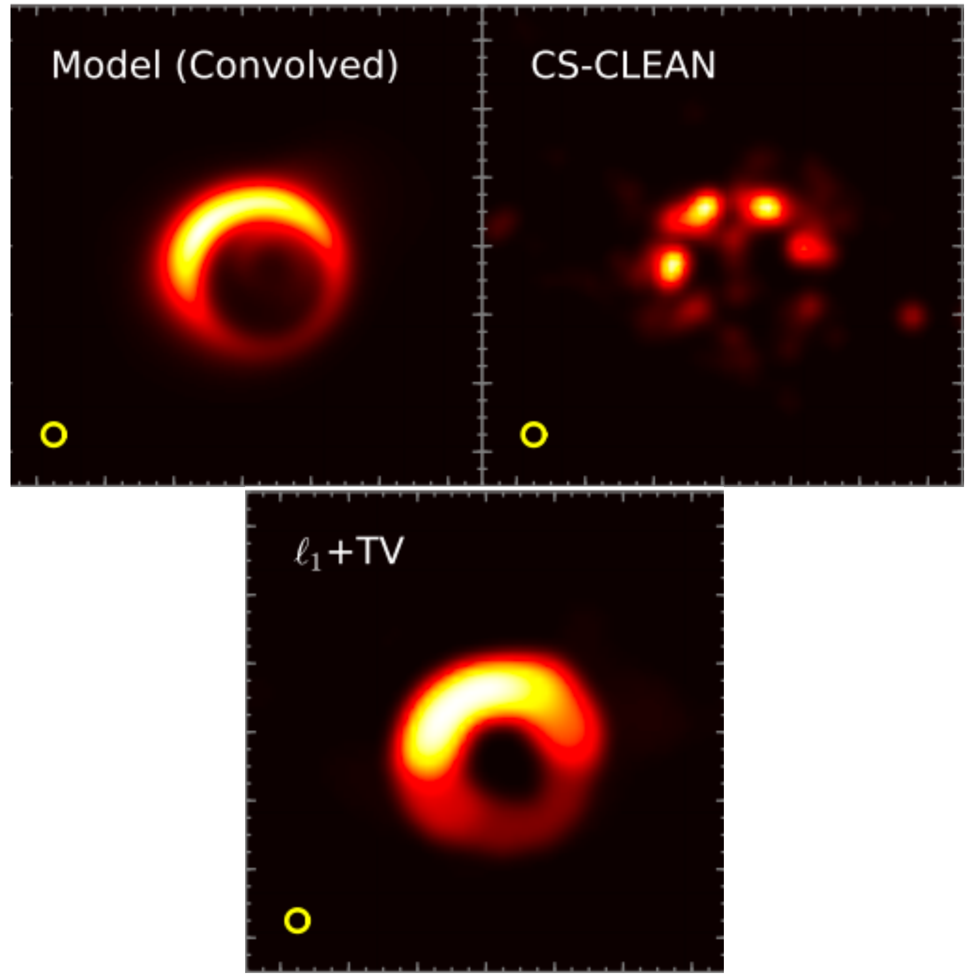


Fig. 5. Simulated UV coverage of M87 with six-station sub-mm VLBI array of EHT. Here it is assumed that observations are conducted at an elevation larger than 20° at each station.

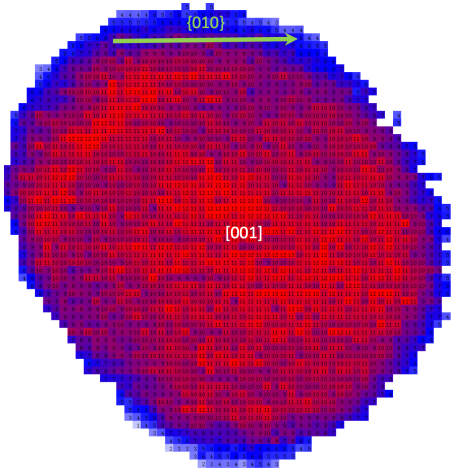


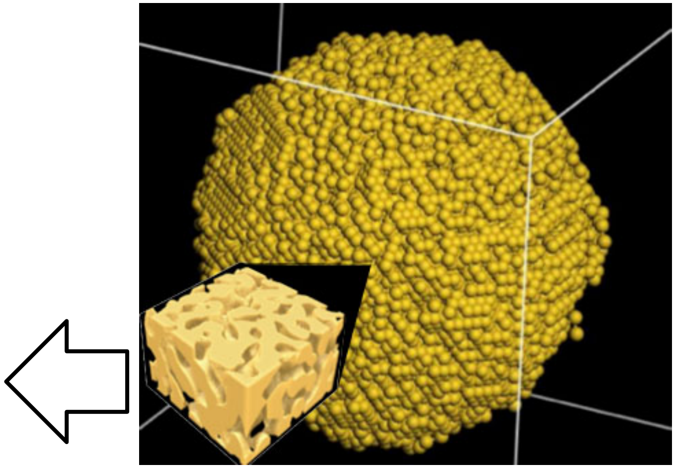


Material and Sparse modeling

Collaboration with C. Nakajima (Tohoku univ.)

- HAADF-STEM
  - Projection mapping





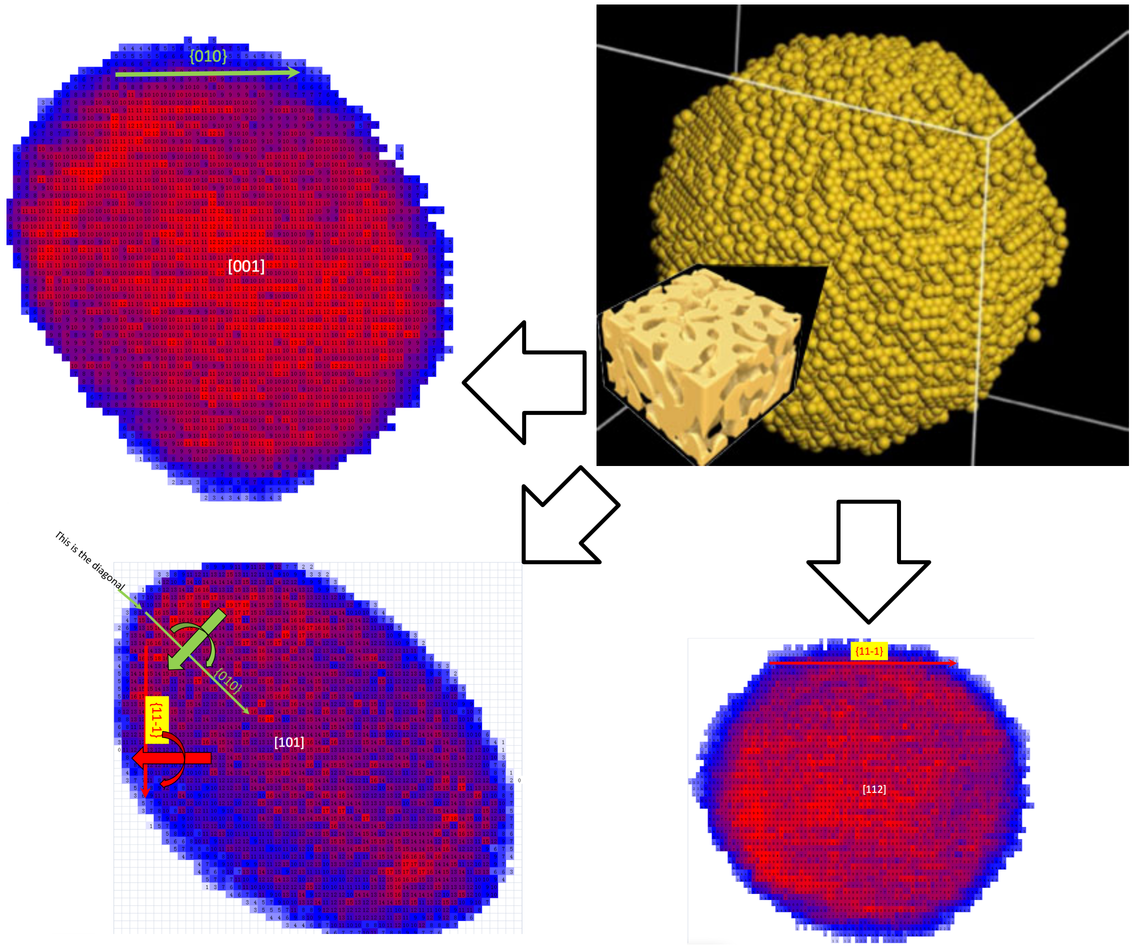
CREST Sparse Modeling ImPACT QUANTUM ANNEALING MACHINE LEARNING

TOHOKU UNIVERSITY

Material and Sparse modeling

Collaboration with C. Nakajima (Tohoku univ.)

- HAADF-STEM
  - Projection mapping

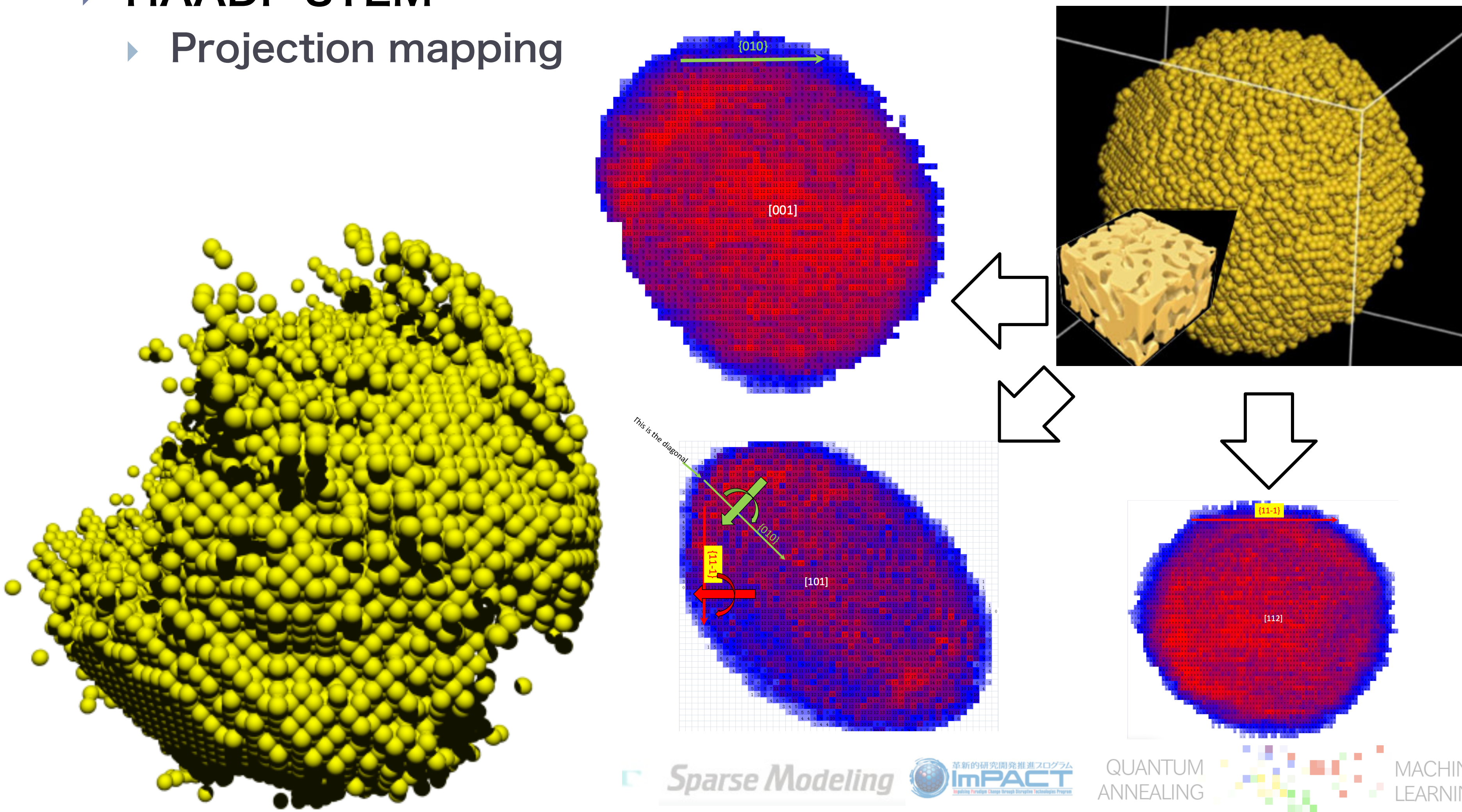


CREST Sparse Modeling ImPACT QUANTUM ANNEALING MACHINE LEARNING

Material and Sparse modeling

Collaboration with C. Nakajima (Tohoku univ.)

- HAADF-STEM
  - Projection mapping

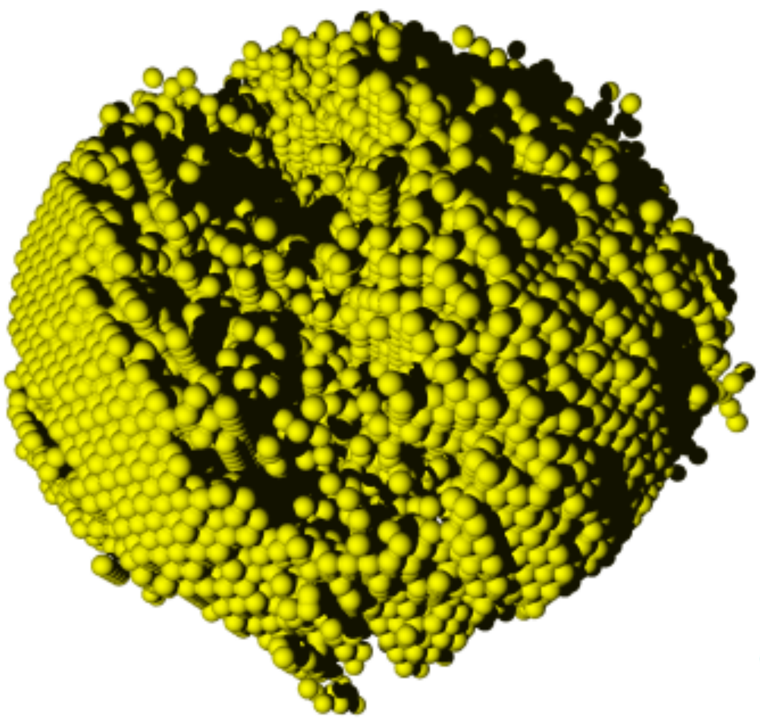


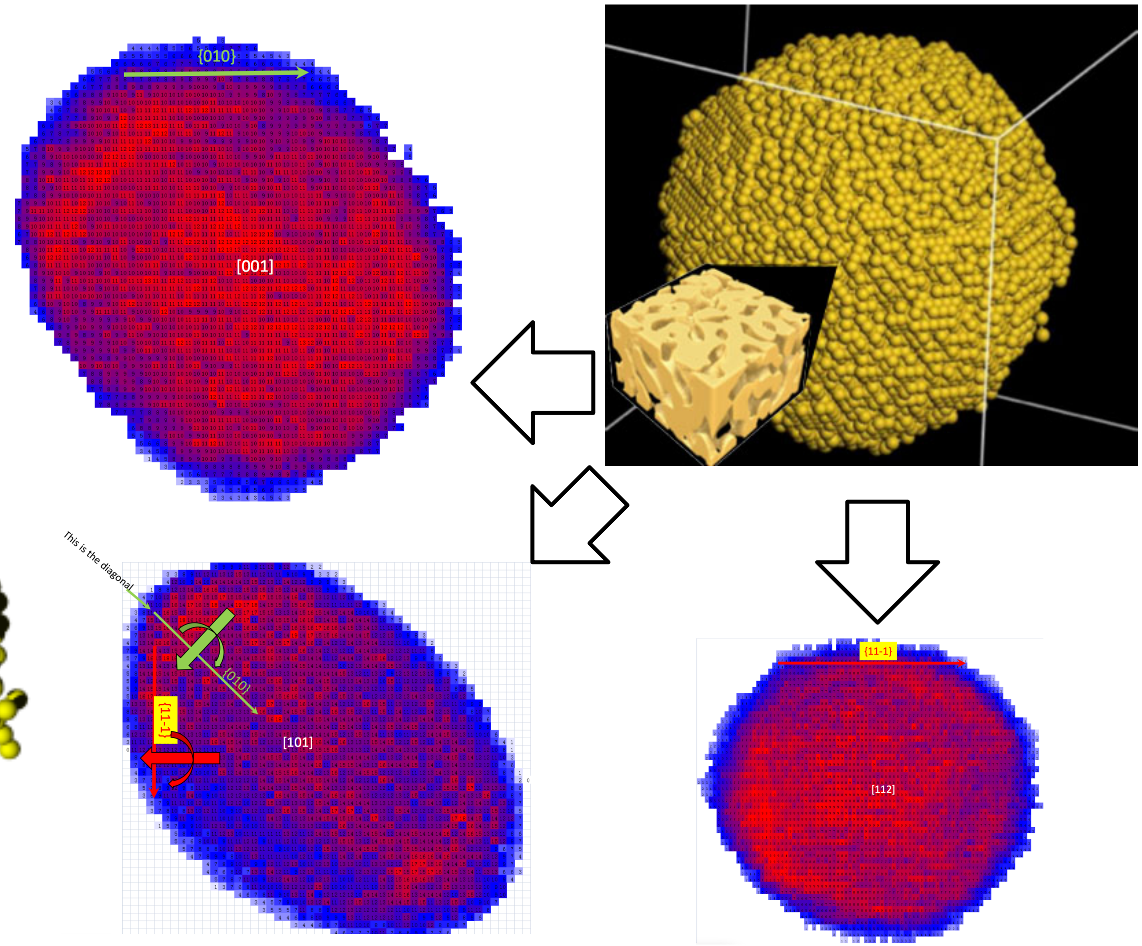
Sparse Modeling ImPACT QUANTUM ANNEALING MACHINE LEARNING

Material and Sparse modeling

Collaboration with C. Nakajima (Tohoku univ.)

- HAADF-STEM
  - Projection mapping



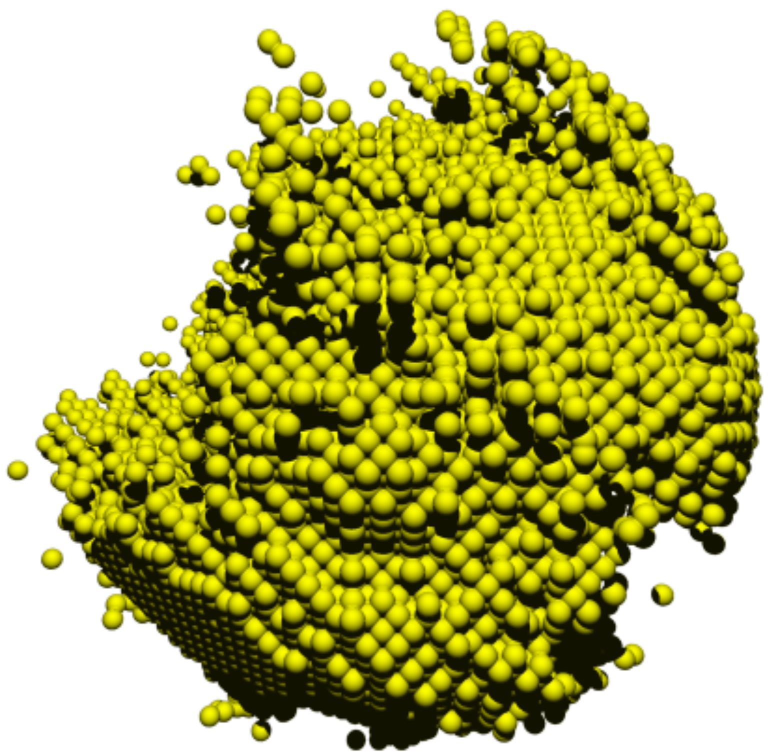


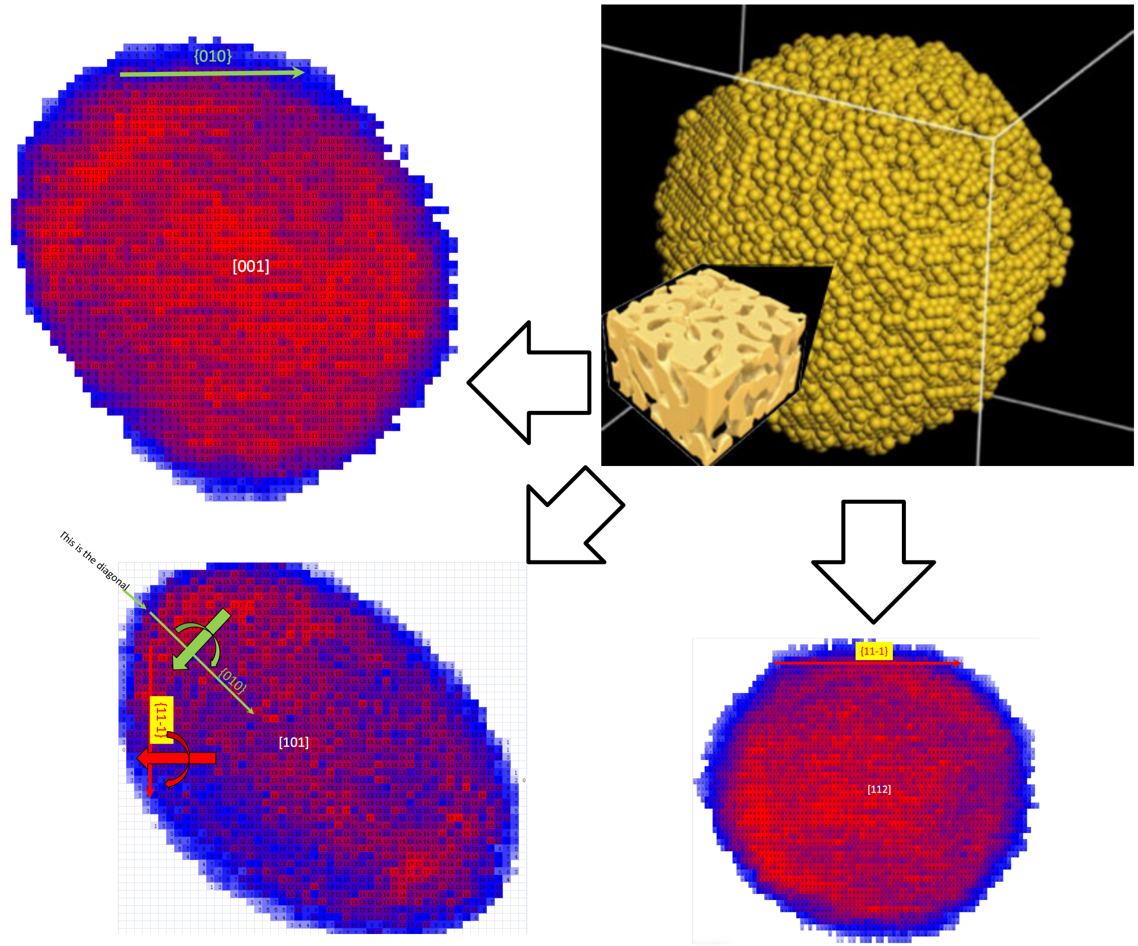
Sparse Modeling ImPACT QUANTUM ANNEALING MACHINE LEARNING

Material and Sparse modeling

Collaboration with C. Nakajima (Tohoku univ.)

- HAADF-STEM
  - Projection mapping



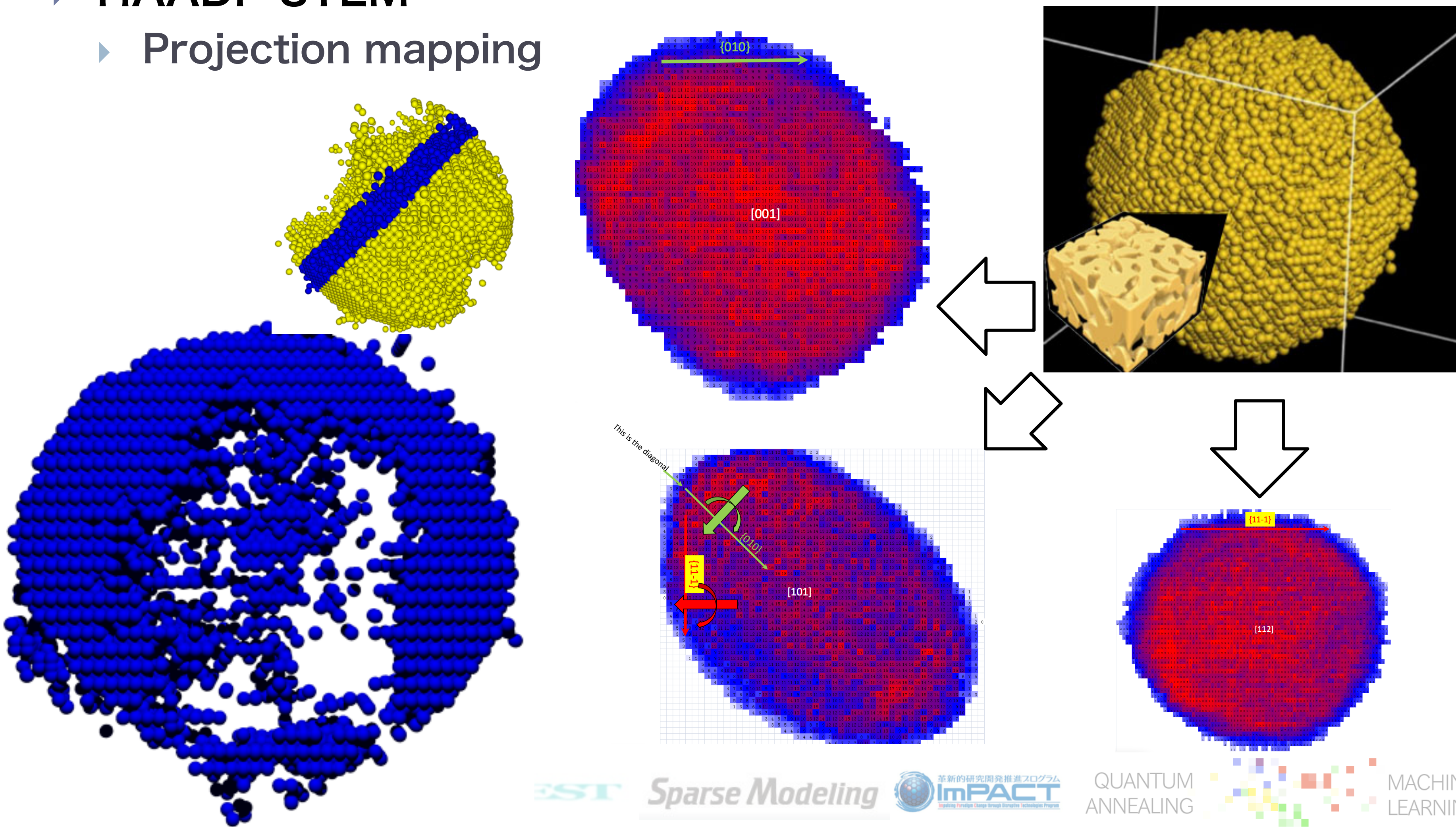


Sparse Modeling ImPACT QUANTUM ANNEALING MACHINE LEARNING

Material and Sparse modeling

Collaboration with C. Nakajima (Tohoku univ.)

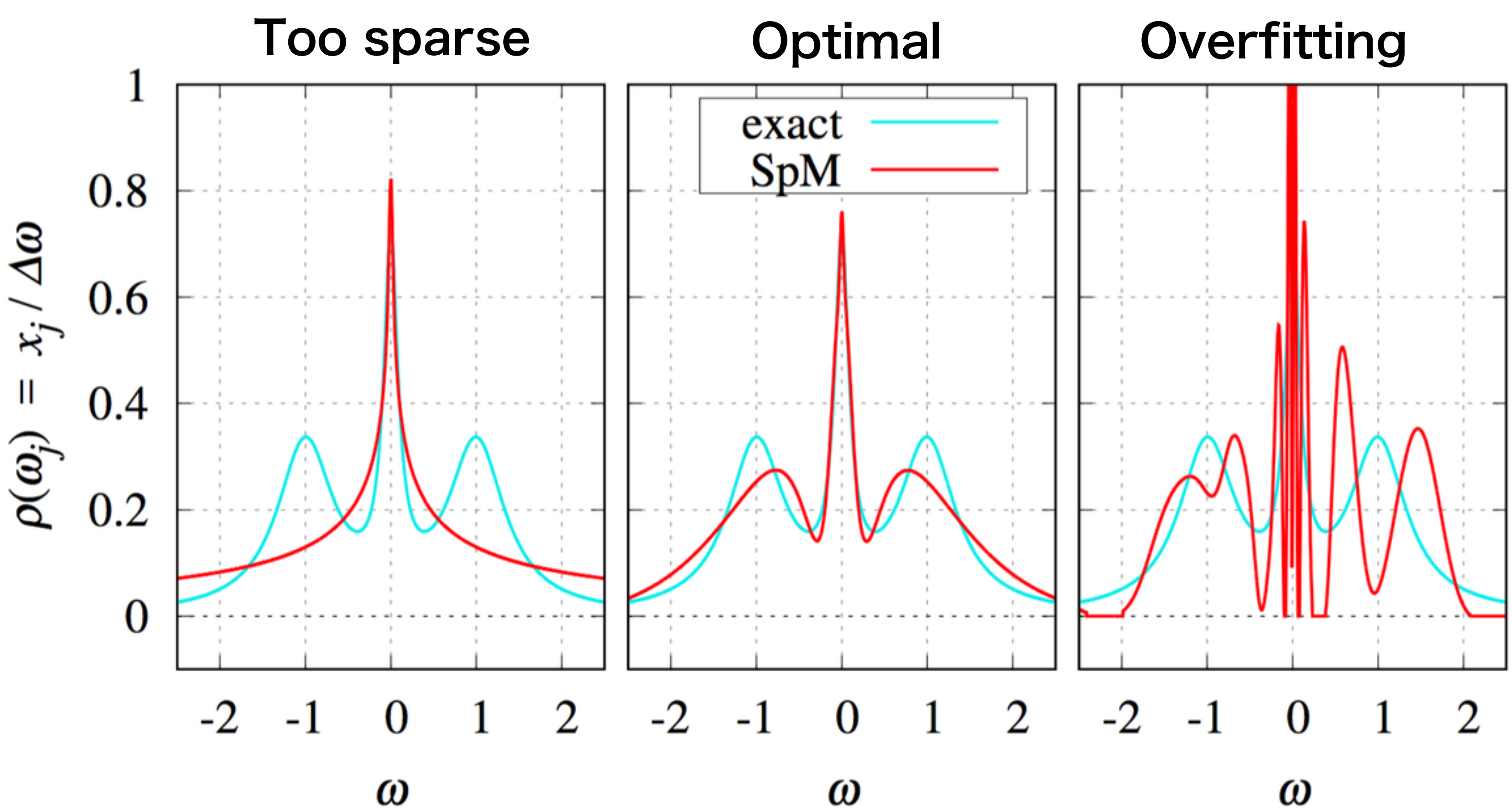
- HAADF-STEM
  - Projection mapping



Analytical continuation in QMC

J. Otsuki, M. Ohzeki, H. Shinaoka, and K, Yoshimi: Phys. Rev. E 95, 061302(R) (2017)

- Results for single-impurity Anderson model
  - Solving $\mathbf{G} = K\boldsymbol{\rho}$ by use of maxEnt? No!!



Lack of information but inference

Compressed sensing

Sparse signal inference

- L0 norm minimization
  - The following optimization problem

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad \text{s.t.} \quad \mathbf{y} = A\mathbf{x}$$

Sparse signal inference

- L0 norm minimization
  - The following optimization problem

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad \text{s.t.} \quad \mathbf{y} = A\mathbf{x}$$

  - L0 norm = number of nonzero elements

Sparse signal inference

- L0 norm minimization
  - The following optimization problem

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad \text{s.t.} \quad \mathbf{y} = A\mathbf{x}$$

  - L0 norm = number of nonzero elements
    - Sparse solution
    - Non-Convex optimization
    - Exponential computational cost (exp(N))

Sparse signal inference

- L1 norm minimization
  - Much easier optimization problem

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{s.t.} \quad \mathbf{y} = A\mathbf{x}$$

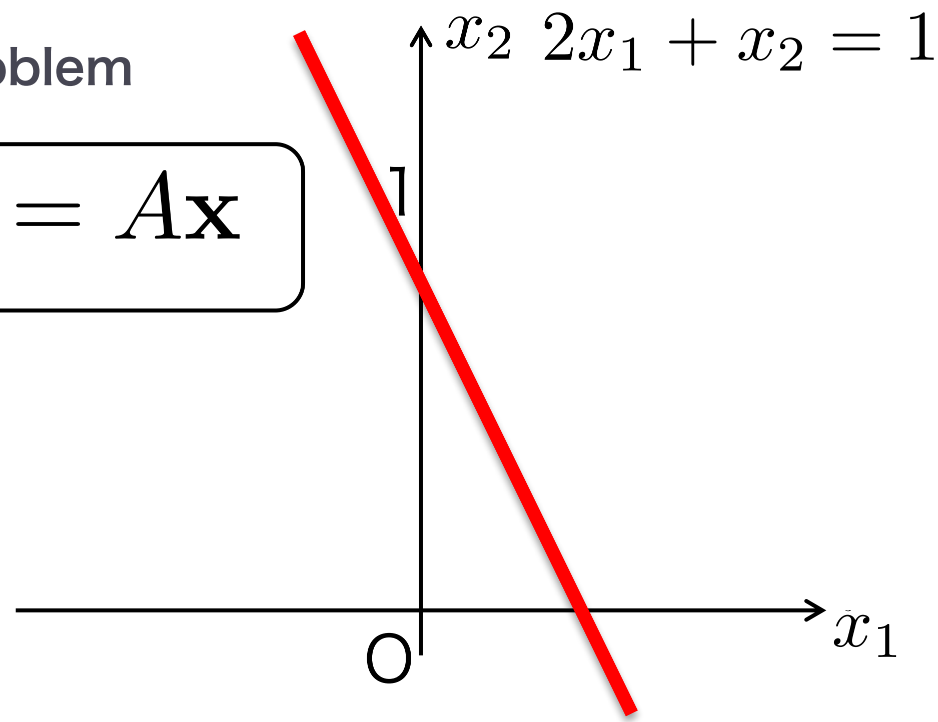
$$\|\mathbf{x}\|_1 = |x_1| + |x_2| + \cdots + |x_N|$$

CREST Sparse Modeling 革新的研究開発推進プログラム ImPACT QUANTUM ANNEALING MACHINE LEARNING

Sparse signal inference

- L1 norm minimization
  - Much easier optimization problem

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{s.t.} \quad \mathbf{y} = A\mathbf{x}$$

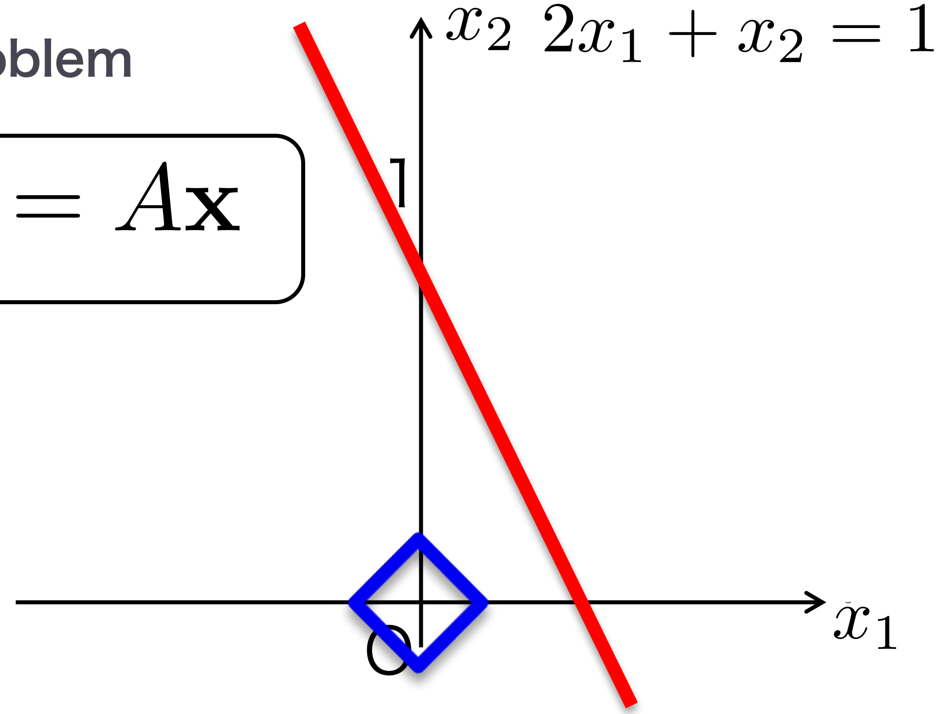


$$\|\mathbf{x}\|_1 = |x_1| + |x_2| + \cdots + |x_N|$$

Sparse signal inference

- L1 norm **minimization**
  - Much easier optimization problem

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{s.t.} \quad \mathbf{y} = A\mathbf{x}$$



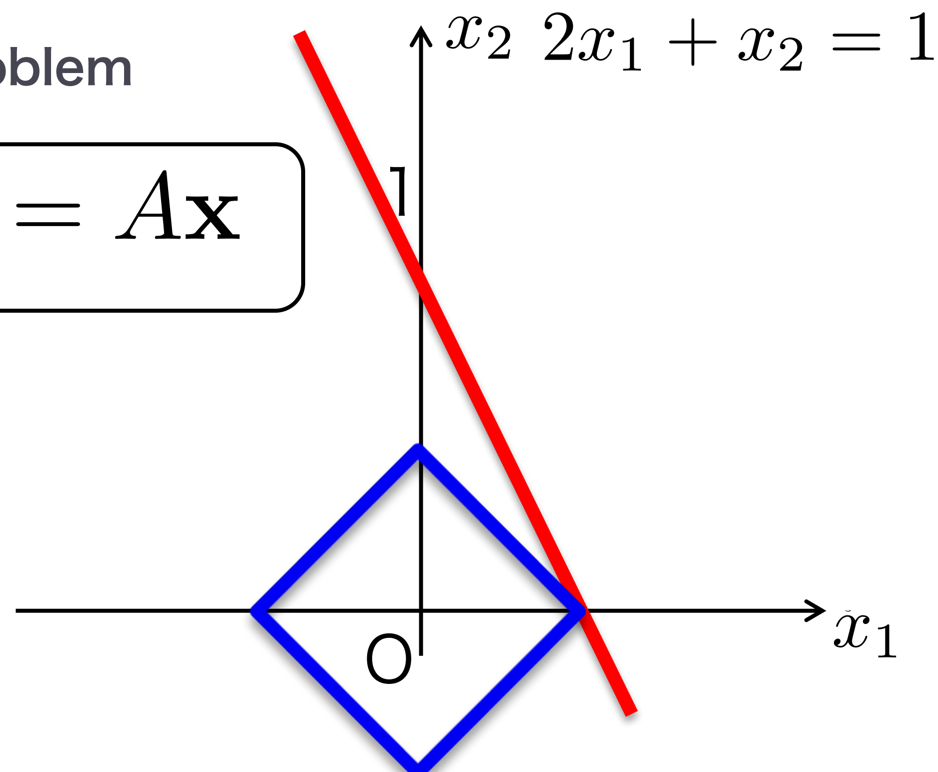
$$\|\mathbf{x}\|_1 = |x_1| + |x_2| + \cdots + |x_N|$$

TOHOKU UNIVERSITY

Sparse signal inference

- L1 norm minimization
  - Much easier optimization problem

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{s.t.} \quad \mathbf{y} = A\mathbf{x}$$

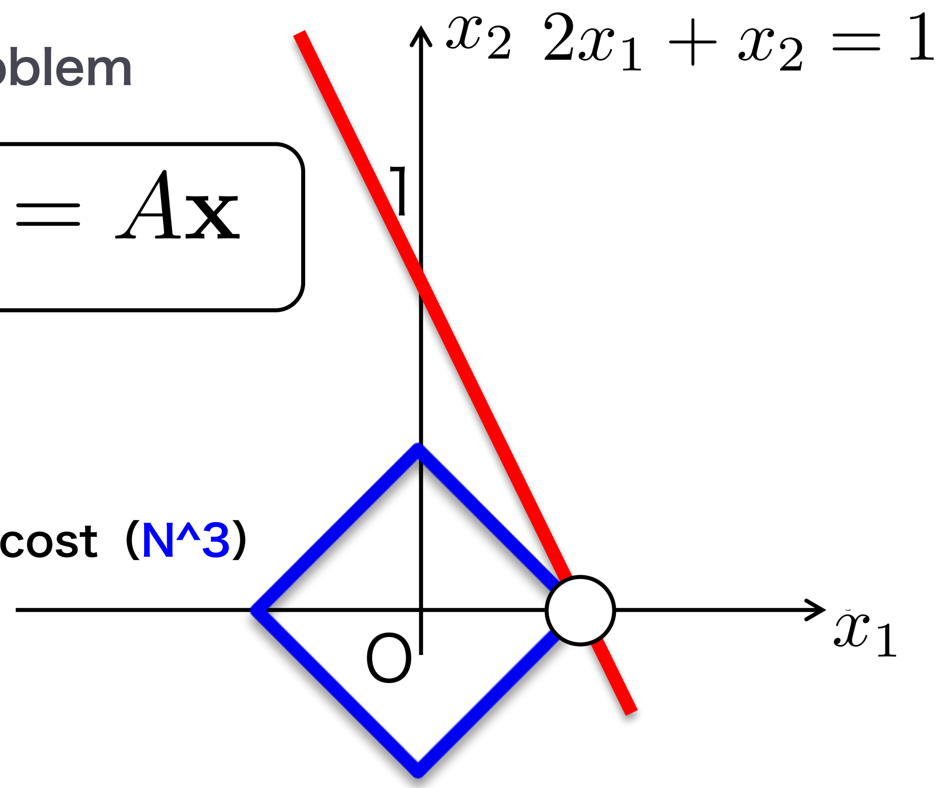


$$\|\mathbf{x}\|_1 = |x_1| + |x_2| + \cdots + |x_N|$$

Sparse signal inference

- L1 norm minimization
  - Much easier optimization problem

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{s.t.} \quad \mathbf{y} = A\mathbf{x}$$



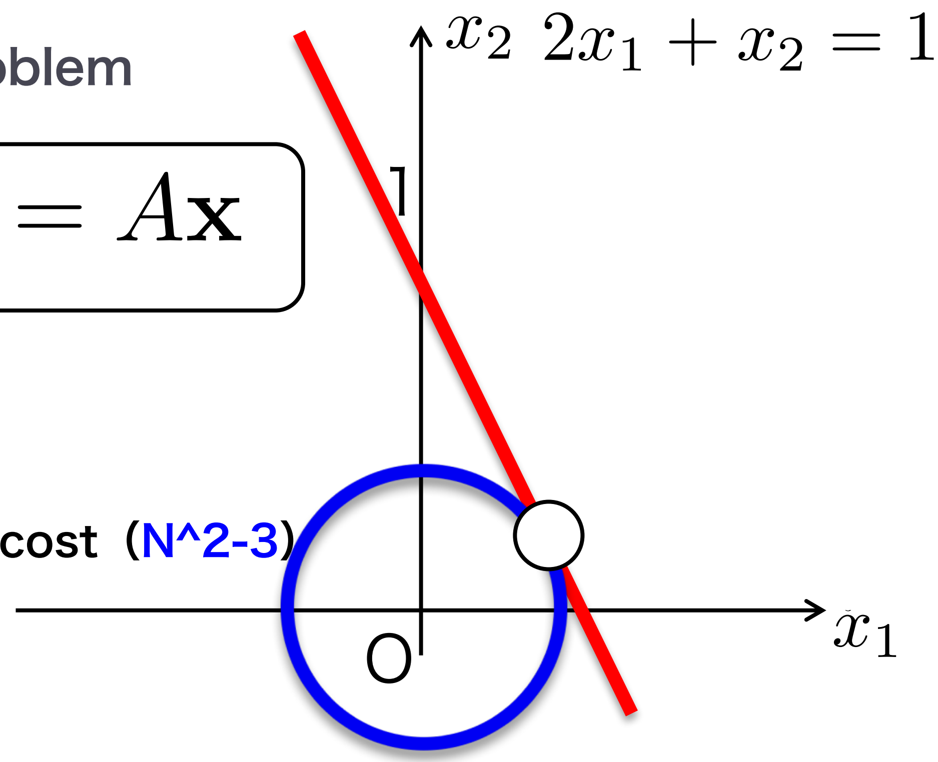
- Sparse solution
- Convex optimization
- Not expensive computational cost (N^3)

$$\|\mathbf{x}\|_1 = |x_1| + |x_2| + \cdots + |x_N|$$

Sparse signal inference

- L2 norm minimization
  - Much easier optimization problem

$$\min_{\mathbf{x}} \|\mathbf{x}\|_2 \quad \text{s.t.} \quad \mathbf{y} = A\mathbf{x}$$



- Non-Sparse solution
- Convex optimization
- Not expensive computational cost (N^2-3)

$$\|\mathbf{x}\|_2 = \sqrt{x_1^2 + x_2^2 + \cdots + x_N^2}$$

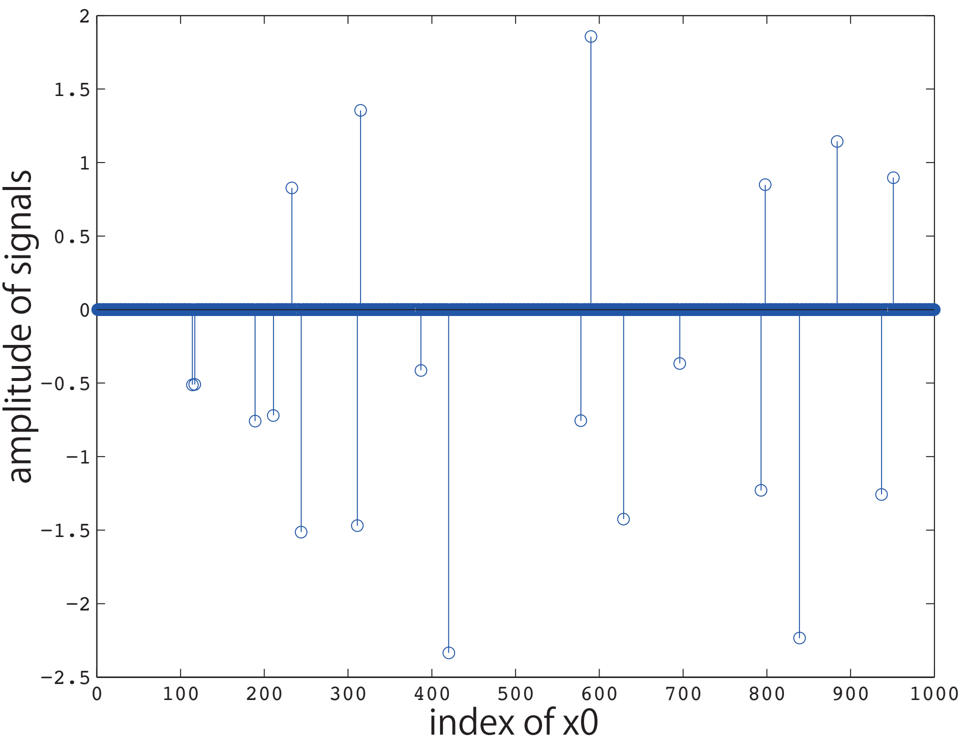
L1 norm selects sparse solution

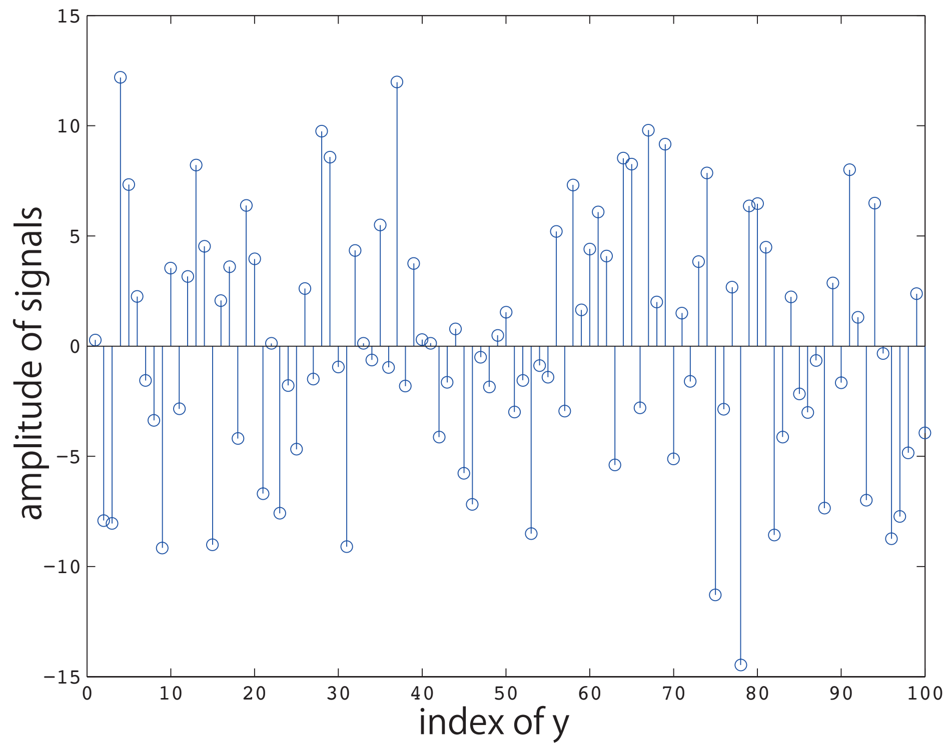
L1 norm selects sparse solution

Correct or not?

Example

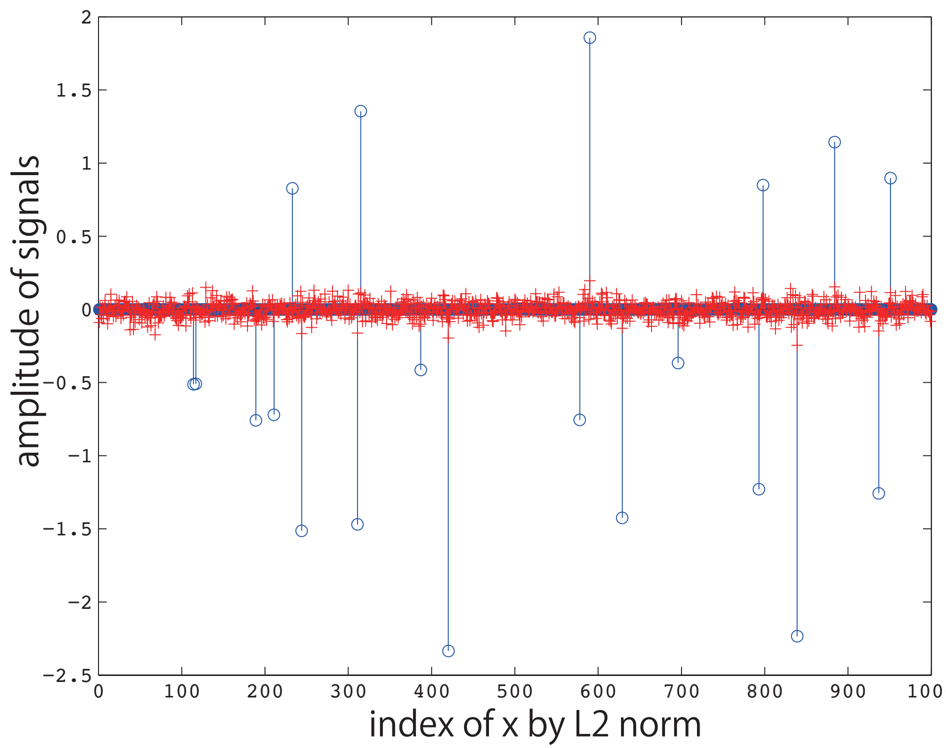
- M=100,N=1000,K=20,A=Gauss random matrix





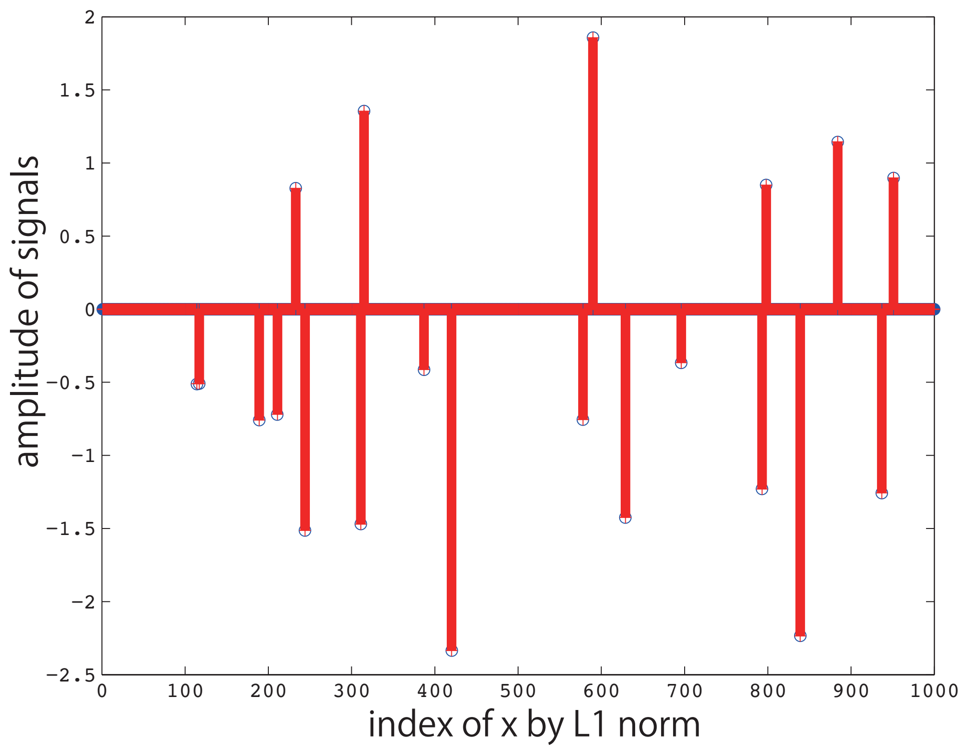
Example

- L2 norm



Example

- L1 norm



L1 norm minimization

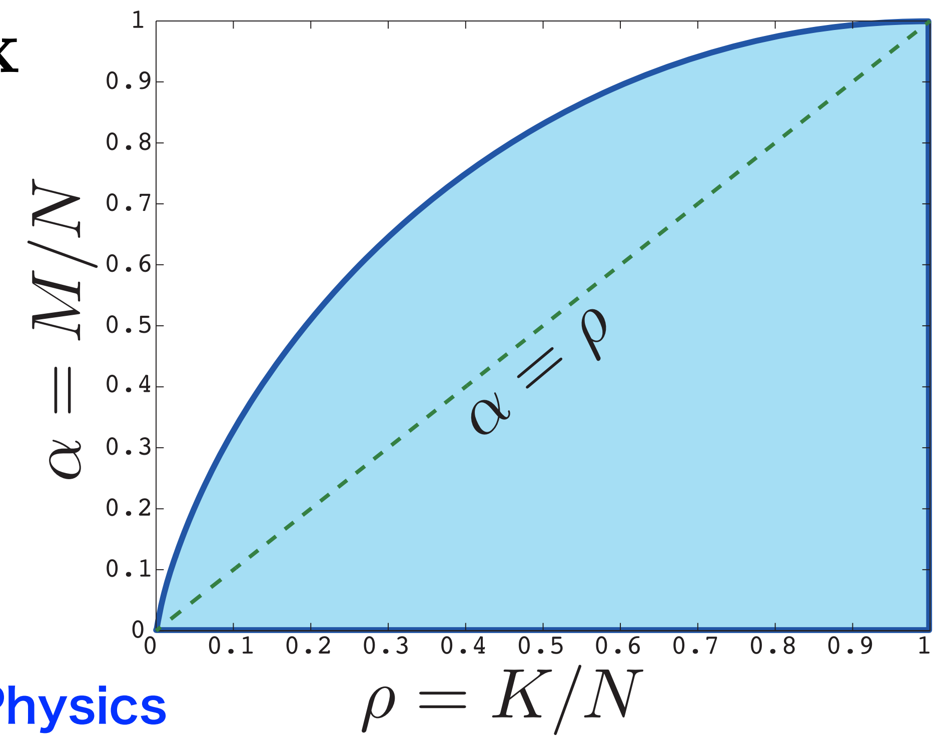
- Performance of L1 norm minimization
  - Prescription：A=Gauss random matrix、x0=Gauss random vector

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{s.t.} \quad \mathbf{y} = A\mathbf{x}$$

$$\frac{1}{\alpha} = 1 + \sqrt{\frac{\pi}{2}} t e^{\frac{t^2}{2}} \{1 - 2Q(t)\}$$

$$\frac{\rho}{1-\rho} = 2\left(\frac{e^{\frac{t^2}{2}}}{t\sqrt{2\pi}} - Q(t)\right)$$

$$Q(t) = \int_t^{\infty} \frac{e^{\frac{x^2}{2}}}{\sqrt{2\pi}} dx$$



- Mathematics or Statistical Physics

[D. L. Dohono and J. Tanner: Proc. Nat. Acad. Sci. 102 (2005) 9452]
[Y. Kabashima, T. Wadayama, and T. Tanaka: J. Stat. Mech.: Theor. and Exp. 09 (2009) L09003]

Sparse modeling

Find x by sparsity

Sparse modeling

Find x by sparsity

Make x sparse

How?

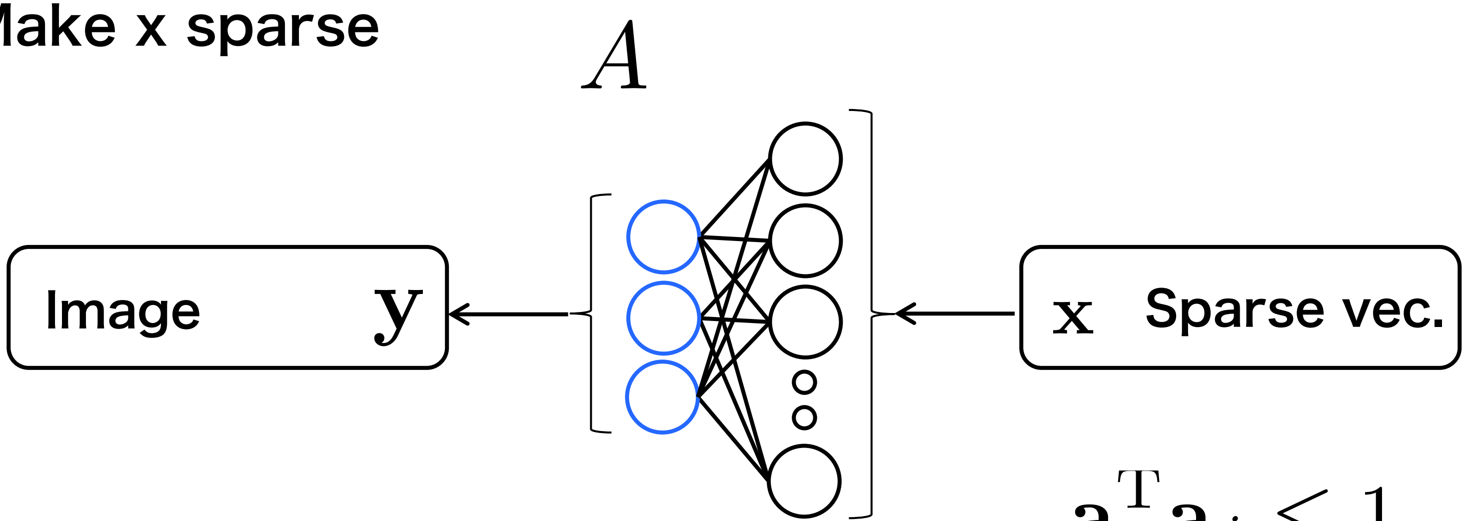
Machine learning!

$$y = f(\mathbf{x})$$

Dictionary learning

J. Mairal, F. Bach, F. Ponce, and G. Sapiro: ICML (2009)

- Make x sparse



$$\mathbf{a}_j^{\mathrm{T}}\mathbf{a}_j \leq 1$$

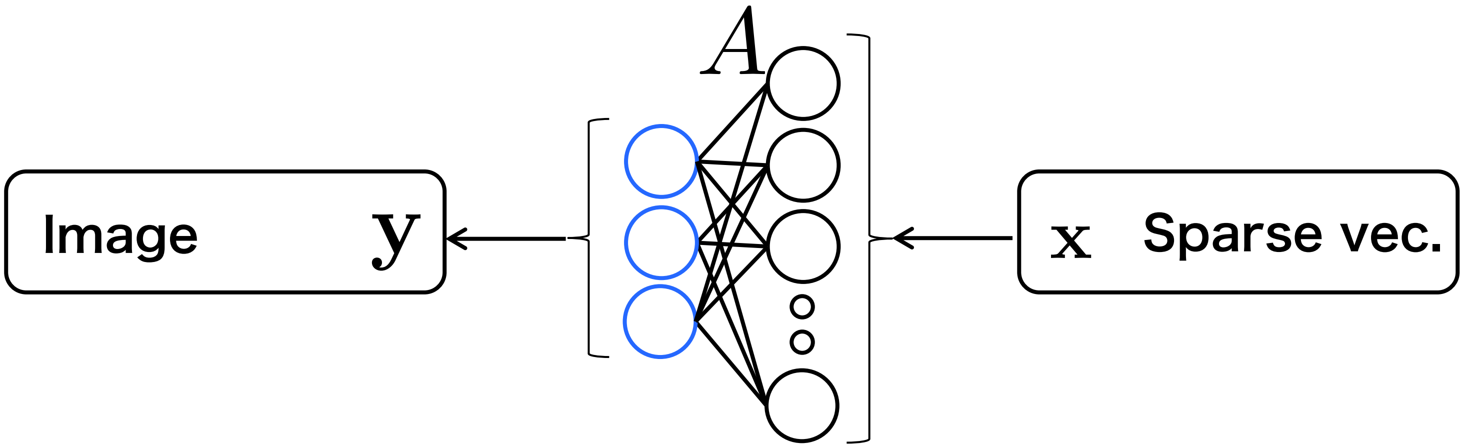
$$\min_{A,\mathbf{x}} \left\{ \frac{1}{2} \|\mathbf{y} - A\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1 \right\}$$

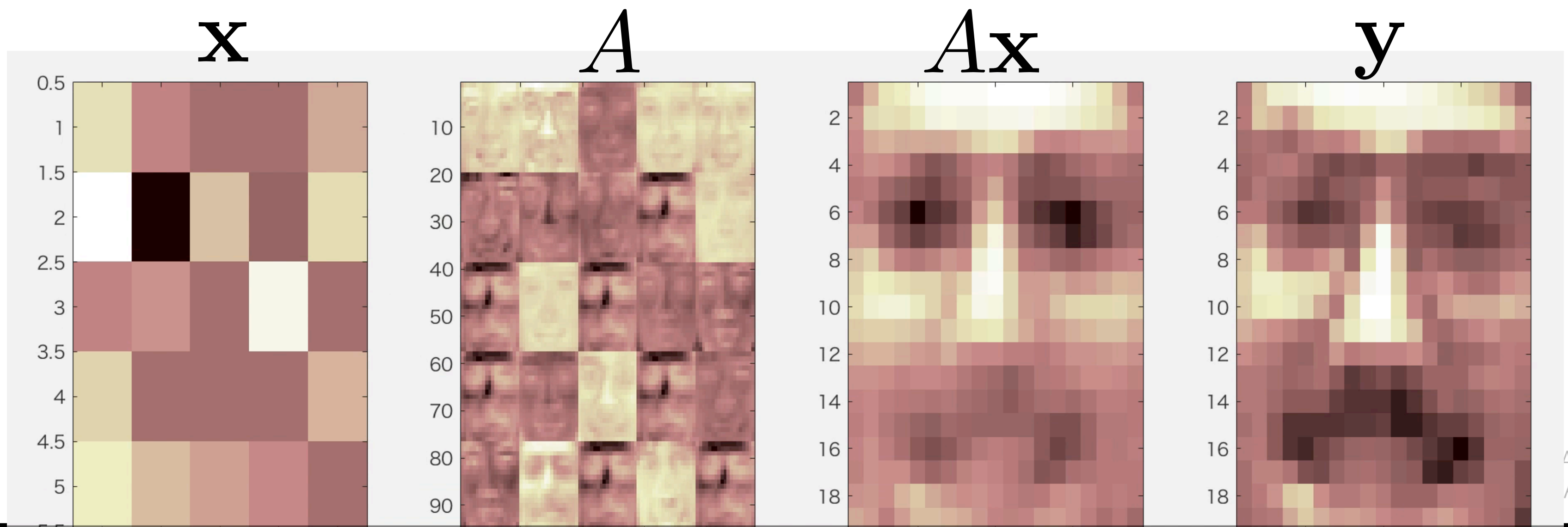
- Find A making x sparse
- For Given y

Dictionary learning

J. Mairal, F. Bach, F. Ponce, and G. Sapiro: ICML (2009)

- Make x sparse





Practical use of Compressed sensing

What is the problem?

- Solve the optimization problem
  - Original problem

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{s.t.} \quad \mathbf{y} = A\mathbf{x}$$

  - Penalty method

$$\min_{\mathbf{x}} \left\{ \frac{1}{2} \|\mathbf{y} - A\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1 \right\}$$

    - LASSO (Least Absolute Shrinkage and Selection Operators)
    - Absolute value ? = Not so difficult!

$$\min_{\mathbf{x}} \left\{ \frac{1}{2} \|\mathbf{x} - \mathbf{v}\|_2^2 + \lambda \|\mathbf{x}\|_1 \right\}$$

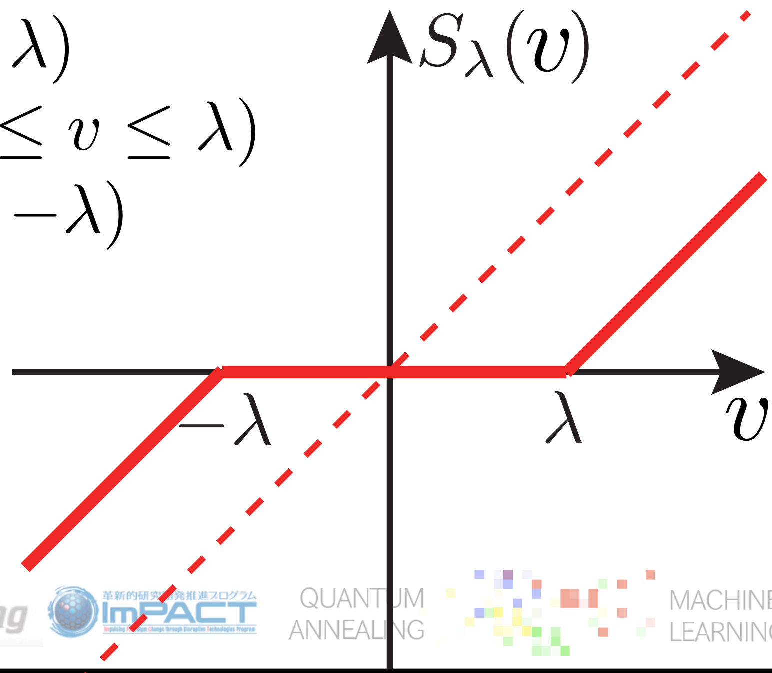
Single-variable problem

- **Soft-threshold function**

$$\min_x \left\{ \frac{1}{2}(x - v)^2 + \lambda |x| \right\}$$

  - Optimal value can be given by

$$x^* = S_\lambda(v) = \begin{cases} v - \lambda & (v > \lambda) \\ 0 & (-\lambda \leq v \leq \lambda) \\ v + \lambda & (v < -\lambda) \end{cases}$$



ADMM [Alternating Direction of Multiplier method)

S. Boyd, et al. Foundation and Trends in Machine Learning, 3 (2010) 1

- Change the problem by augmented Lagrange method
  - Combination of two cost function

$$\min_{\mathbf{x}} \{ f(\mathbf{x}) + g(\mathbf{x}) \}$$

- Ex) LASSO

$$f(\mathbf{x}) = \frac{1}{2} \|\mathbf{y} - A\mathbf{x}\|_2^2 \qquad g(\mathbf{x}) = \lambda \|\mathbf{x}\|_1$$

ADMM [Alternating Direction of Multiplier method)

S. Boyd, et al. Foundation and Trends in Machine Learning, 3 (2010) 1

- Change the problem by augmented Lagrange method
  - Combination of two cost function

$$\min_{\mathbf{x}} \{ f(\mathbf{x}) + g(\mathbf{x}) \}$$

  - Ex) LASSO

$$f(\mathbf{x}) = \frac{1}{2} \|\mathbf{y} - A\mathbf{x}\|_2^2 \qquad g(\mathbf{x}) = \lambda \|\mathbf{x}\|_1$$

  - Splitting

$$\min_{\mathbf{x},\mathbf{z}} \{ f(\mathbf{x}) + g(\mathbf{z}) \} \quad \text{s.t.} \quad \mathbf{x} = \mathbf{z}$$

ADMM [Alternating Direction of Multiplier method)

S. Boyd, et al. Foundation and Trends in Machine Learning, 3 (2010) 1

- Change the problem by augmented Lagrange method
  - Combination of two cost function

$$\min_{\mathbf{x}} \left\{ f(\mathbf{x}) + g(\mathbf{x}) \right\}$$

  - Ex) LASSO

$$f(\mathbf{x}) = \frac{1}{2} \|\mathbf{y} - A\mathbf{x}\|_2^2 \qquad g(\mathbf{x}) = \lambda \|\mathbf{x}\|_1$$

  - Splitting

$$\min_{\mathbf{x},\mathbf{z}} \left\{ f(\mathbf{x}) + g(\mathbf{z}) \right\} \quad \text{s.t.} \quad \mathbf{x} = \mathbf{z}$$

  - Augmented Lagranngian method (multiplier: h、penalty: ρ)

$$\min_{\mathbf{x},\mathbf{z},\mathbf{h}} \left\{ f(\mathbf{x}) + g(\mathbf{z}) + \mathbf{h}^{\mathrm{T}}(\mathbf{x} - \mathbf{z}) + \frac{\rho}{2} \|\mathbf{x} - \mathbf{z}\|_2^2 \right\}$$

ADMM [Alternating Direction of Multiplier method)

S. Boyd, et al. Foundation and Trends in Machine Learning, 3 (2010) 1

- Change the problem by augmented Lagrange method
  - Combination of two cost function

$$\min_{\mathbf{x}} \{ f(\mathbf{x}) + g(\mathbf{x}) \}$$

  - Ex) LASSO

$$f(\mathbf{x}) = \frac{1}{2} \|\mathbf{y} - A\mathbf{x}\|_2^2 \qquad g(\mathbf{x}) = \lambda \|\mathbf{x}\|_1$$

  - Alternation of optimization problem

$$\min_{\mathbf{x}} \left\{ f(\mathbf{x}) + \mathbf{h}^{\mathrm{T}}(\mathbf{x} - \mathbf{z}) + \frac{\rho}{2} \|\mathbf{x} - \mathbf{z}\|_2^2 \right\}$$

$$\min_{\mathbf{z}} \left\{ g(\mathbf{z}) + \mathbf{h}^{\mathrm{T}}(\mathbf{x} - \mathbf{z}) + \frac{\rho}{2} \|\mathbf{x} - \mathbf{z}\|_2^2 \right\}$$

  - Update the multiplier $\mathbf{h} = \mathbf{h} + \rho(\mathbf{x} - \mathbf{z})$

CREST Sparse Modeling ImPACT QUANTUM ANNEALING MACHINE LEARNING

ADMM [Alternating Direction of Multiplier method)

S. Boyd, et al. Foundation and Trends in Machine Learning, 3 (2010) 1

- Change the problem by augmented Lagrange method
  - Combination of two cost function

$$\min_{\mathbf{x}} \{ f(\mathbf{x}) + g(\mathbf{x}) \}$$

  - Ex) LASSO

$$f(\mathbf{x}) = \frac{1}{2} \|\mathbf{y} - A\mathbf{x}\|_2^2 \qquad g(\mathbf{x}) = \lambda \|\mathbf{x}\|_1$$

- Alternation of optimization problem

$$\min_{\mathbf{x}} \left\{ \frac{1}{2} \|\mathbf{y} - A\mathbf{x}\|_2^2 + \mathbf{h}^{\mathrm{T}}(\mathbf{x} - \mathbf{z}) + \frac{\rho}{2} \|\mathbf{x} - \mathbf{z}\|_2^2 \right\}$$

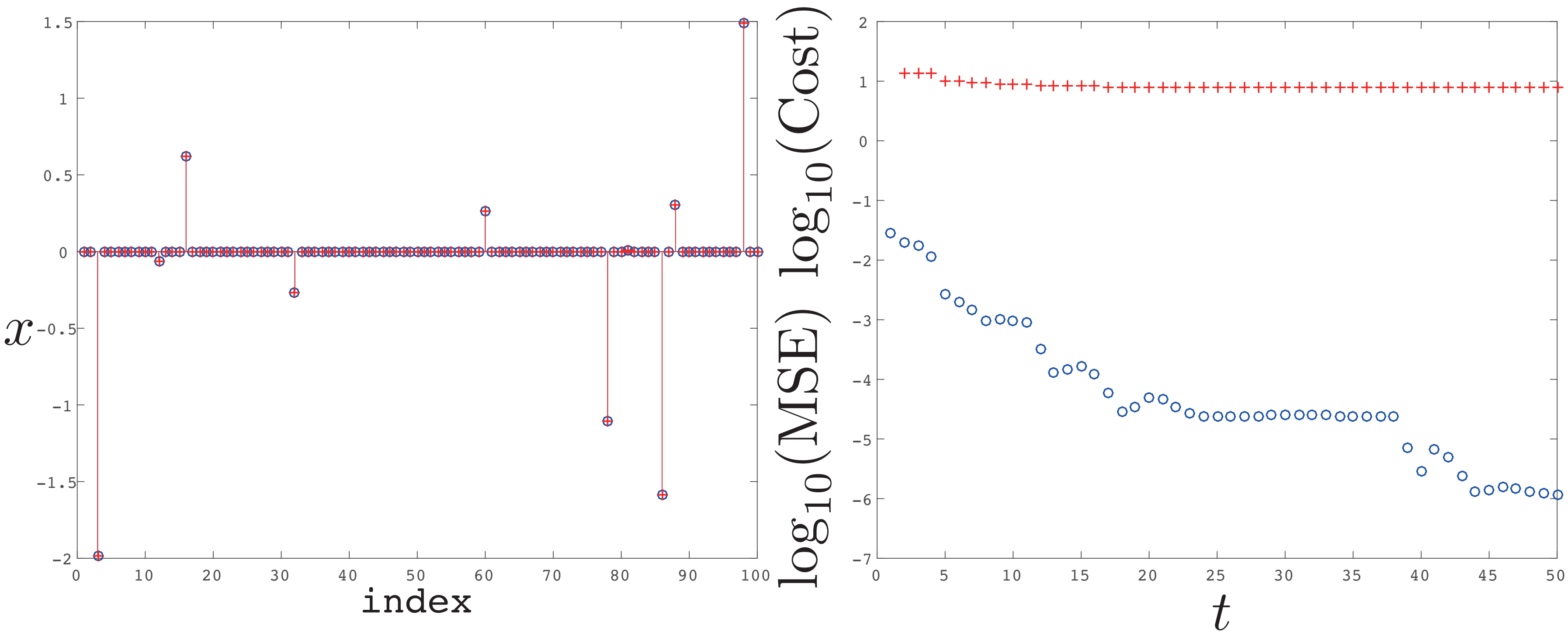
$$\min_{\mathbf{z}} \left\{ \lambda \|\mathbf{z}\|_1 + \mathbf{h}^{\mathrm{T}}(\mathbf{x} - \mathbf{z}) + \frac{\rho}{2} \|\mathbf{x} - \mathbf{z}\|_2^2 \right\}$$

- Update the multiplier $\mathbf{h} = \mathbf{h} + \rho(\mathbf{x} - \mathbf{z})$

ADMM [Alternating Direction of Multiplier method]

S. Boyd, et al. Foundation and Trends in Machine Learning, 3 (2010) 1

- ADMM (given by matlab code in pdf text)



Summary

- Era for Data-driven science
  - Deep learning
    - To identify approximate form of function
    - Renormalization group analysis
  - Sparse modeling
    - To identify most relevant elements from input
    - Extract important structure in nature

  - Application of two modern tools
    - To promote data-driven science
      - Compressed Sensing for recovery from small data
    - To search for new physics
      - Relevant elements from noisy quantum Monte-Carlo data

        J. Otsuki, M. Ohzeki, H. Shinaoka, and K, Yoshimi: Phys. Rev. E 95, 061302(R) (2017)
      - Optimal orthogonal polynomial for analytical continuation

        H. Shinaoka, J. Otsuki, M. Ohzeki, and K, Yoshimi: arxiv:1702.03054

QUANTUM ANNEALING

MACHINE LEARNING